Deep Learning Driven Tool Wear Identification and Remaining Useful Life Prediction

Zhang, Xiaoyang

Award date:
2020

Awarding institution:
Coventry University

Deep Learning Driven Tool Wear Identification and Remaining Useful Life Prediction

By

Xiaoyang Zhang

PhD

September 2020

Deep Learning Driven Tool Wear Identification and Remaining Useful Life Prediction

Xiaoyang Zhang

A thesis submitted in partial fulfilment of the University's requirements for the Degree of Doctor of Philosophy

September 2020

Content removed on data protection grounds

Certificate of Ethical Approval

Applicant:

XiaoYang Zhang

Project Title:

Multi-sensor fusion based online tool condition monitoring and prediction system for milling process

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Low Risk

Date of approval:

19 August 2019

Project Reference Number:

P93854

Content removed on data protection grounds

Acknowledgement

I would express my deep appreciations to those who have help me go through the wonderful journey of my PhD study. To begin with of all, I would thank my director of study, Professor Weidong Li, who has provided me with a valuable opportunity to conduct this doctoral research. He always guides me from his professional perspective on the research and put forward constructive criticism. Meanwhile, he continues to encourage and inspire me during my study until I put this work into practice.

I would also thank my research supervisors, Dr Xin Lu and Dr Sheng Wang. They have shared valuable opinions and inspired me to keep on this research. Their consistent trust and supervision are an integral part of this research.

Also, I want to thank my parents and family. They are layman for the research, and they support me unconditionally from all aspects. Their continuing inspiring, encouragement and company allow me to complete my doctorate study firmly.

Finally, I want to thank everyone who assisted me during this study. Their friendliness also the bolster to me. Words are powerless to express my gratitude to them, and their great support and help burned in my mind!

Table of Contents

List of figures

List of tables

Abbreviations

1D	1-Dimensional
2D	2-Dimensional
2D-DCT	Two-dimensional Discrete Cosine Transform
AC	Alternating Current
AE	Acoustic Emission
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
AV	Average Value
CC	Cloud Computing
CNN	Convolutional Neural Network
CV	Cross Validation
CWT	Continuous Wavelet Transform
DBN	Deep Belief Network
DC	Direct Current
DE-SVR	Differential Evolution SVR
DFT	Discrete Fourier Transform
DNN	Deep Artificial Neural Network
CPU	Central Processing Unit
EC	Edge Computing
EKF	Extended Kalman Filter
EMD	Empirical Mode Decompositions
EM-PCA	Expectation-maximisation for Principal Component Analysis
ET	Extra Trees
F	Force
FDR	Fisher's Discriminant Ratio
FFT	Fast Fourier transform
GASF	Gramian Angular Summation Fields
HMM	Hidden Markov Model
IoT	Internet of Things
ISOMAP	Isometric Mapping
KNN	K-nearest neighbours
LLE	Local Linear Embedding

LR	Logistic Regression
LSTM	Long Short-term Memory
MAD	Median Absolute Deviation
MAE	Mean Absolute Error
MDS	Multidimensional Scaling
MLP	Multilayer Perceptron
mRMR	Minimum redundancy maximum Relevance
NB	Naive Bayes
NFN	Neural-fuzzy Network
PCA	Principal Component Analysis
PCC	Pearson correlation coefficient
PHT	Perceptual Hashing Technology
PSD	Power Spectral Density
PSO	Particle Swarm Optimisation
ReLU	Rectified Linear Unit
RF	Random Forest
RFE	Recursive Feature Elimination
RMS	Root Mean Square
RNN	Recurrent Neural Network
RUL	Remaining Useful Life
SBULSTM	Stacked Bi-directional and Uni-directional LSTM
SLGT	Supervised Learning of Genetic Tracking
SMEs	Small and Medium-sized Enterprises
SVM	Support Vector Machine
SVR	Support Vector Regression
TCM	Tool condition monitoring
TP	Turning Point
TSBP	Trajectory Similarity-based Prediction
V	Vibration
WPT	Wavelet Packet Transform
WSN	Wireless Sensor Network

Abstract

Tooling condition monitoring and prognosis has been researched and developed extensively considering its importance of supporting manufacturing systems. An effective system for tooling condition monitoring and prognosis can ensure good product quality, minimisation of tooling failure and optimisation of production cost. However, due to the complex mechanism in manufacturing processes, the tooling condition depends on different factors (i.e., machining parameters and tooling/product materials). Therefore, the conventional experiment-based methods are expensive, time-consuming, and less effective. In recent years, the industrial Internet of Things (IoT) related technologies and the state-the-art artificial intelligence algorithms (e.g., deep learning algorithms) have been increasingly applied to manufacturing applications for mining valuable information from massive amounts of data, thus making complex manufacturing processes easier to be understood and improved. Therefore, in this thesis, based on multiple sensors and deep learning algorithms, a series of novel and systematic methods for tooling condition monitoring and prognosis for a machining system are proposed. The aim of this research is to enhance the overall performance of tooling condition monitoring and prognosis in terms of computational efficiency, prediction accuracy and system robustness.

This research covers the two primary indicators for tool condition monitoring and prognosis, that is, tool wear identification and tool remaining useful life (RUL) prediction. For tool wear identification, in this research, based on multi-sensor signals, a new method, i.e., recursive feature elimination and cross-validation (RFECV), is designed to optimise feature selection and fusion. In the process of RFECV, a support vector machine (SVM)-based classifier is used to recursively evaluate the contributions of different feature subsets to the classification of the tool wear for optimal feature selection. Dimensionality reduction on the selected features is further implemented via the isometric mapping (Isomap) method. Finally, a concise and robust 1D convolutional neural network (CNN) model is devised to perform tool wear identification using the newly generated features from the Isomap.

To improve accuracy and expedite computational efficiency for predicting the RUL of cutting tools for a machining system, a novel methodology is designed. The methodology integrates strategies of signal partition and deep learning algorithms for effectively processing and analysing multi-sourced sensor signals collected throughout the lifecycle of a cutting tool. In more detail, the methodology consists of two subsystems: (i) a Hurst exponent-based method is developed to effectively partition multi-sourced signals along with the tool wear evolution; (ii) a hybrid CNN-LSTM (convolutional neural networks-long short-term memory) algorithm is developed to

combine feature extraction, fusion and regression in a systematic means to facilitate the prediction based on segmented signals.

In order to meet the needs of small and medium-sized enterprises (SMEs), in this work, an economical, three-level, and multi-sensor-based monitoring and prognosis system is designed. In the system, a wireless acquisition platform collects multi-sensor signals. To expedite computational efficiency and minimise data traffic, an edge computing (EC)-based gateway is embedded as an extended Kalman filter for signal denoising, performing the conversion of the time-series data format to reduce the high latency of transmission. A hybrid CNN-RF (random forest) model is devised to achieve real-time tool wear identifications. Furthermore, based on the received signal image, the system can conduct RUL prediction based on an integrated multi-channel CNN-LSTM model at the cloud computing end.

Two open-source experimental datasets and a workshop deployment case were used to verify the practicability and reliability of the methods and systems proposed in this research. Regard to the method for tool wear identification, it was verified that the prediction of the selected feature subsets achieved the best prediction accuracy. Moreover, the proposed RFECV-SVM approach was proved to be superior to other machine learning models on feature selection. For the tool RUL prediction, the data partition method was developed for assigning the signal data corresponding to each tool wear stage, and case studies were used to compare the designed hybrid deep learning algorithm with some other main-stream algorithms, such as CNN, LSTM, DNN (deep artificial neural network) and PCA (principal component analysis), under the conditions of partitioned and un-partitioned signals. The conducted performance comparison showed that, the methods proposed in this research is essentially better than those of the comparative algorithms. In addition, the developed EC-enabled tool prognosis system was validated that it can provide a comprehensive and satisfactory diagnosis result, in terms of tool wear identification and tool RUL prediction. The EC effectively reduced the amount of data, improved transmission efficiency and enhance the data privacy of the conventional IoT system. The proposed system was proved to be effective, flexible and affordable.

Chapter 1. Introduction

1.1 Research background

To leverage the rapid technical progresses of information technology, automation and artificial intelligence, some major economies in the world have proposed new industrial strategies, such as Germany's Industry 4.0 (Kagermann et al. 2013), Made in China 2025 (State Council of China 2015) and USA Advances Manufacturing Partnership (Rafael et al. 2014), to transfer existing manufacturing paradigms to more automated and agile manufacturing, across the entire layers of factories.

Monitoring and control of manufacturing processes is a crucial task for smart manufacturing. In the process, the status of machining cutting tools is considered an essential factor to decide workpiece quality, tooling life, productivity and energy consumption in manufacturing. According to Zhou and Xue (2018), 7-20% downtime of the CNC machine is caused by tool failure, 3-12% total cost of entire machining processes are wasted on replacing the cutting tools, and there are only 50-80% cutting tool life has been effectively used. Early detection of tool anomalies can significantly help reduce frequent downtime for tool changes, cut down costs, and ensure the quality of products. The tool wear identification and tool remaining useful life (RUL) prediction are both vital research areas that have been actively investigated in recent years (Xu et al. 2018).

So far, two approaches have been widely used to implement the wear status identification and RUL prediction of a cutting tool, that is, experience-based and data-based. An experience-based model is a traditional method that requires to record a large amount of machining cycles in order to summarise the law of tool wear (Benkedjouh et al. 2013). Nevertheless, it is difficult to build up a mathematical model for cutting tool monitoring in real-time. As a well-established tool wear evaluation theory, the Tayler's equation shows that, processing parameters, such as cutting speed and feed rate influences the cutting tool status to a large extent during machining processes (McParland et al. 2016). In reality, apart from the above factors, a machining process unavoidably involves many other factors, which affect the tool condition stochastically. Some of these factors are difficult to be measured directly.

Thanks to the advances in information technologies, data-based models for tool wear monitoring and prognosis have been increasingly researched in recent years. The models have demonstrated great potentials in supporting applications by providing the more dynamic and effective interpretation of a cutting tool's status based on condition monitoring signals. A research carried out by the Kennametal Company in the United States (Zhou and Xue 2018) showed that

a CNC system provided with a tool condition monitoring (TCM) system could lay aside up to 30% of the machining cost caused by insufficient use of the cutting tool, and avoid low surface quality and dimensional errors of workpieces that due to an excessively worn cutting tool. However, to better use data-based models, research needs to be conducted further.

In practical, multiple sensors offset the shortage of a single signal on the tool status monitoring. However, the massive data collected from different sensors deployed on a machine tool system may contain duplicate information and noises, which will hugely prolong the processing time and increase the system complexity. Therefore, it is prominent to design a data-based system to leverage the rich information from multi-sensors as well as identifying unwanted data for removal, thereby enhancing the processing efficiency and increasing the accuracy of decision-making.

To develop an efficient monitoring and prognosis system for cutting tool conditions, research should not just concentrate on sensor signal processing, but a suitable reasoning algorithm is also crucial. Commonly used algorithms for tool conditions are based on machine learning algorithms, such as fuzzy logic (Rajamani et al. 2018), support vector machine (SVM, Pandiyan et al. 2018), artificial neural networks (ANN, Shankar et al. 2019) and so on. On the other hand, although such a machine learning algorithm provides a computing environment for the construction of the relationship between sensor signals and variables for decision making (e.g., classification, regression, clustering), it still faces the challenge to mining hidden information and features from data (Schmidt et al. 2019). With the quick progress of deep learning algorithms and their successful applications in various engineering problems, intelligent algorithms on tool conditions should no longer be limited to the traditional machine learning algorithms. It is imperative to explore how to effectively use deep learning algorithms, such as the deep neural network (DNN), convolutional neural network (CNN) and long short-term memory (LSTM), to support tool condition monitoring-based prognosis.

Furthermore, it also needs to design an appropriate IoT (internet of things) system to integrate the above technologies/models and achieve an optimal system performance. A traditional IoT system for TCM means to collect condition monitoring data through sensors, and then synchronise at a cloud computing (CC) centre (Motlagh et al. 2020). There are some commercial TCM systems available on the market (e.g., systems from Montronix, Nordmann, Prometec and Marposs) (Sheng 2012). And the systems of Brankamp and Kistler claimed to have the tool condition monitoring capabilities to some extends (Mandal 2014). However, the primary function of these systems is for data acquisition rather than tool wear identification or tool RUL prediction. Moreover, the cost of implementing these systems is expensive and less customisable for SMEs. Meanwhile, cloud centres are popularly used for data processing and storage in the system, making data transmission heavy (Khan et al. 2017). Edge computing (EC) has been recently

introduced to complement a cloud enabled IoT system to balance local and centric processing. Edge computing refers to a distributed network edge server that is additionally deployed on the basis of the traditional cloud computing architecture. It obtains corresponding data from the data terminal and realises the downstream processing of the cloud server with specific performance and computing resources (Lin et al. 2020, Zhang et al. 2021). It is promising to provide a timely response at the local end, decrease the amount of data delivered to the cloud platform and improve information safety and privacy (Chen and Ran 2019). Thus, it is necessary to design an edge-cloud system to integrate monitoring and prognosis methods to achieve integrated functions and overall system optimisation.

1.2 Research challenges

In view of the above industrial needs, related research gaps still exist. In this work, investigations will be carried out from the following aspects.

Sensor signal fusion. To enhance the accuracy of cutting tool monitoring and prognosis, multi-sensor based methods are in active development (Arcady et al. 2018). To facilitate decision making, data fusion technologies have been investigated to handle various signals and eliminate redundant signals. Data fusion could be categorised as data-level fusion, feature-level fusion and decision-level fusion (Beddar-Wiesing and Bieshaar 2020). However, raw sensor signals usually contain plenty of unorganised information collected during signal acquisition processes. Although there are some methods have been designed for raw signal denoising and pre-processing, for instance, wavelet transform (Mitiche et al. 2017), Welch's method (Ferguson et al. 2018), low-pass filter (Zhou et al. 2019), correlated information and features in correspondence with a cutting tool status are still difficult to extract from the raw data based on these methods (Aguileta et al. 2019). Thus, to address the above issues, in this research, a suitable fusion strategy for tool status monitoring and prognosis based on multi-sensor signals will be developed.

Optimal feature selection. If there are plenty of features extracted from multi-sensor signals not associated with tool wear, it is difficult to design an appropriate algorithm for accurate classification or prediction (prognosis). Feature selection has been explored by many researchers in order to reduce feature dimensions and discard redundant features in some applications (Mares et al. 2016, Kaddar et al. 2019 and Chen et al. 2019). However, in the tool condition monitoring related fields, the relevant research is rare. Meanwhile, it is challenging to reveal the relationship between extracted features and the cutting tool status by using existing feature selection methods, such as KPCA (kernel principal component analysis), LLE (local linear embedding) and mRMR

(minimum redundancy maximum relevance). Therefore, in this work, the relevant investigation will be conducted.

Tool RUL prediction. Besides the tool wear identification, the prediction of tool RUL through regression analysis is also of great significance to improve production cost-saving and productivity. However, this technology is still in infancy as only a few of research on tool life prediction has been carried out. The cutting tool RUL prediction relies on the prediction of the flank wear during an entire life cycle of machining, which is usually partitioned into three regions: initial wear, stable wear and severe wear. Different tool wear conditions and stages will lead to the imbalanced distribution of collected signals for corresponding stages, which may bias the prediction results and increase prediction errors (Zhang et al. 2020). To minimise the inaccurate prediction in this situation, divide the signals into reasonable fragments between each wear stages will produce a promoting effect. Though some works have been carried out in the aspect (Zong et al. 2013; Qian et al. 2014; Rodriguez et al. 2014; Ali et al. 2015; Wang et al. 2016; Guo et al. 2016; Leevy et al. 2018; Mao et al. 2018; Low et al. 2018; Opałka et al. 2018), none of these proposed approaches provides effective signal partition method, which should be researched further in this work.

Intelligent algorithm design. Conventionally, the ANN, fuzzy inference system or SVM have been employed to support tool status monitoring, while their performances are not satisfactory in processing massive data. Active research works have been undertaken to develop intelligent algorithms for tool status monitoring and prognosis using machine learning algorithms (Dutta et al. 2016; Patra et al. 2017; Corne et al. 2017; Cuka and Kim 2017; Yuan et al. 2018; Kong et al. 2019; Cao et al. 2019). In the last few years, deep learning algorithms have been applied as they are capable to effectively and automatically draw features from datasets, and CNN and LSTM are ones of the most popular algorithms (Song et al. 2017; Núñez et al. 2018; Fawaz et al. 2019; Kong et al. 2019, Zang et al. 2020; Zhu et al. 2020; Zhang et al. 2020). To leverage the merits of the algorithms and overcome their weakness, hybrid deep learning algorithms are in an active investigation. Due to the complexity of cutting tool prognosis and the industrial requirements for high accuracy and efficiency, it is necessary to design effective hybrid deep learning algorithms to provide a reliable prediction.

IoT system architecture design. In order to promote the execution of Industry 4.0, manufactories have integrating cutting-edge information technologies in the production lines to sustain the competitiveness in the world market. With the advance of IoT technology and cloud centres, some low-cost cloud server enabled IoT-based TCM systems were developed in the laboratories (e.g. Mourtzis et al. 2016, Caggiano 2018). However, with that the number of sensors

attached to the facilities has remarkably increased, the amount of collected data has been exponentially growing, which leads to the massive growth of transmission bandwidths, data transmission power consumption, storage space and computation resource requirements for cloud servers (Tabassam 2017). That is, the current cloud-enabled IoT architecture cannot sufficiently meet industrial requirements, in terms of the bandwidth insufficiency, inevitable network delay, the high expense of big data transmission and lack of data security mechanism. To address the issue, edge computing-enabled data processing frameworks have been actively designed and applied, also including the areas of machine condition monitoring (Qiao et al. 2020; Lou et al. 2020; Setz and Aiello 2020; Gia et al. 2020; Rasheed et al. 2020). To leverage the advantage of edge computing, in this research, edge computing will be integrated cloud computing as an IoT architecture to better tool condition monitoring and prognosis.

1.3 Research aims and objectives

The overall goal of the work exhibited in this thesis is to develop a tool status monitoring and prognosis system based on multi-sensor signals and deep learning algorithms, for achieving satisfactory accuracy and computational efficiency on the tool condition monitoring to meet industrial especially SMEs' needs. The aim will be achieved by addressing the following objectives:

- To conduct a thorough literature survey in the field of tool condition monitoring, classification and prognosis, which contains the multi-sensor signal fusion technology, signal partition, deep learning algorithms and edge computing-based monitoring system.
- To develop optimal feature selection and fusion methods to optimise and synchronise features based on signals under different domains, and to improve the accuracy and computational efficiency of tool wear identification.
- To develop an effective strategy to segment a large volume of signals to facilitate the deep learning-based signal processing, thereby balancing the signal feature distribution in various tool wear stages, and further improving the performance of the deep learning algorithms in tool condition monitoring and prognosis.
- To design appropriate hybrid deep learning algorithms to fuse signals and their inherent features effectively in supporting more accurate prediction on cutting tool conditions, by leveraging the strengths of individual deep learning algorithms.
- To develop an affordable and flexible multi-sensor monitoring system that is configured with edge/cloud-based architectures for wider applicability, and integrate the above methods and algorithms to provide a more precise and dependable result for tool condition monitoring, identification and prognosis.

1.4 Research datasets

In order to verify the performance of the proposed systems in term of the tool wear identification, the tool RUL prediction, and the portable IoT monitoring prognosis, datasets that acquired from actual manufacturing experiments were employed. Two of the datasets are from the publicly available experimental dataset: NASA milling dataset (Agogino and Goebel 2007) and PHM Society Conference Data Challenge (PHMSociety 2010), and another was collected from an industrial workshop. The details of these datasets are given in later chapters. In addition, relevant comparative analysis was conducted to evaluate the advancement of the proposed technologies. On the other hand, the processing and analysis of signal data, the prediction and decision-making of the results are all based on powerful software that conducted on personal computers, including MATLAB, Apache Spark (python programming language based), TensorFlow and so on.

1.5 Research contribution

This thesis focuses on the aim and objectives of the study and obtains the following contributions.

1. Presented a thorough review of up-to-date developments in tool status monitoring, classification and prognosis-related technologies.
2. Developed a novel feature fusion system that integrates the function of sensor signal denoising, optimal feature selection and feature dimensionality minimisation, to obtain effective feature subset with the optimal number of features for different signals under complex cutting operations, thereby maintaining the minimal correlation between features, and promoting the effectiveness of tool wear identification with a lower overfitting risk.
3. Designed a 1D-CNN model to build an accurate data-driven prediction model for processing features in an efficient means.
4. Developed a novel systematic methodology to integrate strategies of signal partition based on Hurst exponent and hybrid CNN-LSTM algorithms, for effectively processing and analysing multi-sourced sensor signals throughout the lifecycle of a cutting tool.
5. Devised a portable and cost-effective edge computing-enabled wireless system to detect tool wear and predict tool life with customisable configuration architectures.
6. Validated the proposed systems based on different datasets including those from actual industrial cases, to confirm that the systems are effective in identifying tool wear status and predicting tool RUL in satisfactory accuracy and computational efficiency.

7. Conducted detailed comparisons between the proposed systems and other main-stream algorithms to demonstrate the superior performance of the research.

1.6 Publication arising from the thesis

Zhang, X.Y., Lu, X., Wang, S., Wang, W., and Li, W.D. (2018) "A Multi-Sensor Based Online Tool Condition Monitoring System for Milling Process". Procedia CIRP 72, 1136–1141, Sweden.

Zhang, X.Y., Lu, X., Li, W.D., and Wang, S. (2020) "Prediction of the Remaining Useful Life of a Cutting Tool Using the Hurst Exponent and CNN-LSTM". The International Journal of Advanced Manufacturing Technology. (accepted)

Zhang, X.Y., Lu, X., Li, W.D., and Wang, S. (2020) "Prediction and Optimisation for Remaining Useful Life of a Cutting Tool". Accepted as a chapter for Data Driven Smart Manufacturing Technologies and Applications (Springer).

Zhang, X.Y., Lu, X., Li, W.D., and Wang, S. (2020) "Feature Selection and Optimisation for Tool Wear Identidication". The Journal of Computers & Industrial Engineering (Submitted).

Zhang, X.Y., Lu, X., Li, W.D., and Wang, S. (2021) "An Edge Computing Based Architecture for Tool condition prognosis". The International Conference of Manufacturing Research, Derby.

1.7 Thesis structure

The above content introduces the research background, motivation and purpose of this research. The remaining chapter of the thesis is organised as follows.

Chapter 2 investigates the literature referring to the tool status diagnosis, as well as some related concepts and theories, including the tool condition monitoring system, existing technologies of data processing, partition and data fusion, the up-to-date application of deep learning algorithms in the field of machinery monitoring and the development of data acquisition platform.

Chapter 3 proposes a method for tool wear classification based on multi-sensor fusion and the 1D CNN model. The designs of signal pre-processing, feature extraction, feature selection, feature fusion, and the CNN model, are presented. Comparisons and analyses are given.

Chapter 4 proposes a tool RUL prediction method based on a series of signal processing and a hybrid CNN-LSTM model. Moreover, the performance evaluation by comparing the proposed method with other competitive algorithms are carried out.

Chapter 5 constructs an affordable multi-sensor IoT monitoring system for the tool status prognosis. The architecture design and integration with intelligent algorithms to support decision making on tool conditions are detailed. Benchmark assessment with other prevalent methods is presented.

Chapter 6 summarises the deliverables of the entire research and discusses the potential future possibilities of the study.

Chapter 2. Literature review

2.1 Introduction

The high precision and the low consumption manufacturing have motivated researchers to investigate the improvement of the tool condition prognosis system regards the tool wear identification and the tool RUL prediction. In particular, the effective sensor signal processing, data fusion and the application of deep learning algorithms have achieved initial results in the field of machinery monitoring. Moreover, along with the increasing development of intelligent IoT technology, research related to the robust and flexible monitoring system has also shown its necessity, in the aspect of sensor monitoring system architecture, diagnostic algorithm, etc.

In this chapter, the survey of sensor signal processing technologies, data fusion strategies and deep learning algorithm of the previous tool status prognosis research is presented. Meanwhile, the sensor monitoring platform development and the EC application also included. Together they constitute the primary basis and starting point of this work.

2.2 Tool condition monitoring (TCM)

In the manufacturing industry, the cutting tool failure will increase the maintenance cost and time, even reduces productivity. TCM systems have become a promising approach to reduce the occurrence of unwanted situations significantly. Balsamo et al. (2016) investigated the catastrophic tool failure detection through the multi-sensor monitoring system, the signal data of vibration and AE (acoustic emission) sensor was segmented into the different samples, then the features of each sample were extracted to indicate the tool fails. The presented system could remind the tool failure, but the absence of the feature selection and the decision-making model led to the detection delay and low accuracy. Rizal et al. (2017) proposed a tool wear detection system with multi-sensor signals, in terms of cutting force, vibration and temperature. The signal feature was input into the Mahalanobis-Taguchi system to implement the tool wear classification and detection, after the feature extraction in time- and frequency-domain. The experiment validation achieves 88% detection accuracy, however, the correlation between each feature and their contribution to the tool wear is not considered.

With the growing potential of multi-sensor signals in tool status monitoring, most studies have realised that the signal feature reflects more excellent value than the original signals, and the selection of the extracted feature could achieve higher monitoring performance. Caggiano (2018) have collected the cutting force, vibration and AE signal during the cutting process of Ti alloy.

After extracting the signal feature, he applied the PCA to estimate the reasonable signal feature for dimensionality reduction. Finally, the proposed tool status prediction system diagnosed the tool condition through ANN. By adopting the ANN with PCA on different sensor signals, the selected feature in this system is helpful for tool status monitoring and tool replacement. However, the feature in the different signal domain is lack of consideration, which leads to information insufficiency. Moreover, the PCA is unable to reveal valuable information of the nonlinear sensor signal, the redundant signal amount in the subsequent processing is bound to increase. Besides, Krishnakumar et al. (2018) developed a feature-level fusion classification system of tool status, and the decision tree has been adopted for the dominant feature selection and feature fusion of AE and vibration signal. Then, the decision tree, SVM, ANN and NB (naive Bayes) assessed both the fused and non-fused feature subsets, and the combination of fused feature subset and ANN model produces the highest classification efficiency. The signal fusion method they have proposed shows the effectiveness of the tool status monitoring, whereas, the optimal feature selection is not conducted based on different signals, thus, the prediction performance is limited.

CNN attracts many considerations, because of its efficient and precise performance. Although CNN often used in image processing field, it has emerged in recent research for tool status prediction system. Terrazas et al. (2018) captured the force signal of dry milling for the tool status prediction, these signals were converted to 2D image format by Gramian angular summation fields (GASF), which is symmetry matrix, then directly trained with CNN model. Aghazadeh et al. (2018) developed a tool status recognition system with a CNN model, the feature of force and vibration sensor signal was extracted by wavelet transform, and the training result of the Bayesian network, SVM, KNN (K-nearest neighbours) regression model and CNN model showed that the CNN obtained the highest performance. Gouarir et al. (2018) explored the tool wear prediction with three-axis vibration signals, and a CNN-based prediction system was proposed. The reconstructed image format data of the original signal directly input into the 2D CNN model without signal denoising and feature selection. Finally, only a general prediction accuracy was achieved. Huang et al. (2019) introduced a CNN prediction system integrated with multi-sensor fusion. The statistical features of the signal were extracted from the time-domain, frequency-domain and time-frequency-domain. The verification of force and vibration signal shows that the performance of the CNN model is significantly superior to the SVM model. The above studies are summarised in Table 1.

Table 1. The literature of tool condition monitoring

Sensor Signal	Feature selection	Prediction model	Reference
AE, vibration	×	Signal segmentation	Balsamo et al. (2016)
Cutting force, vibration, temperature	×	Mahalanobis-Taguchi	Rizal et al. (2017)
Cutting force, AE, vibration	PCA	ANN	Caggiano (2018)
AE, vibration	Decision tree	ANN	Krishnakumar et al. (2018)
Force	GASF	2D CNN	Terrazas et al. (2018)
Force, vibration	wavelet transform	2D CNN	Aghazadeh et al. (2018)
Vibration	×	2D CNN	Gouarir et al. (2018)
Force, vibration	×	2D CNN	Huang et al. (2019)

2.3 Cutting tool RUL prediction

Flank wear occurs on the surface of the cutting tool that contact with the workpiece and caused by abrasive wear. As the most common form of tool wear, flank wear can provide a more reliable RUL estimation (Zhang et al. 2016). In addition, the cutting tool life is not an absolute concept but depends on the choice of different tool life indicators. The life of cutting tools can usually be expressed by the following indicators (Jacso, Matyasi and Szalay 2019):

- Cutting tool effective machining time.
- The number of parts processed during tool operation.
- The total length of the cutting path of the tool during tool operation.
- The total surface area of the workpiece processed during tool operation.

Moreover, many inconsistent factors affect and restrict the life of cutting tools. For example, in the finishing process, surface quality and dimensional accuracy are the most important, and in the roughing process, excessive cutting force and vibration are the limiting factors (Yousefi and Zohoor 2019). Therefore, it is not enough to evaluate tool life based on the existing empirical formula, which is too subjective and ignores the actual environmental factors during machining.

In recent years, many studies of the monitoring system that incorporated with sensor data have focused on the establishment of a cutting tool RUL model.

Karam et al. (2016) studied the multi-sensor online tool RUL assessment method based on signal processing, feature extraction and ANN. The cutting force, vibration and AE signals were processed and analysed for feature extraction, these features then were input into ANN for tool life prediction. Zhang et al. (2016) collected three-axis vibration signals during milling and selected their critical features according to the Pearson correlation coefficient (PCC). Then, adopted neural-fuzzy network (NFN) to conduct the prediction of tool wear and the RUL. Nevertheless, this study only focuses on a single sensor, the accuracy is not guaranteed. Yu et al. (2017) proposed a new method of tool wear diagnosis based on hidden Markov model (HMM). The root mean square (RMS) of the vibration signal was used as a health indicator to construct multiple HMM in a weighted manner for the prediction, and the tool wear has been divided into discrete wear region by HMM for better observe the tool wear evolution. Experimental results show that the weighted HMM provides satisfactory prediction of the RUL. However, a single signal source and the complex model become obstacles to practical application. Wu et al. (2018) proposed a cutting tool RUL prediction system based on multi-sensor fusion. The statistical features of vibration, AE and current signals were extracted from time and frequency-domain. In consideration of the relationship between time-varying signal and the cutting tool wear, these features were manually segmented, then, the correlation analysis, monotonicity analysis and residual analysis methods were adopted for feature selection. Next, an adaptive network-based fuzzy inference system (ANFIS) was used for achieving feature fusion, finally, the RUL prediction was conducted by polynomial curve fitting method. The results display that the introduced scheme is feasible and has good prediction performance. However, the generalisation of the proposed model is limited because parameters in the prediction model are set based on experience. Yang et al. (2019) established a prediction model of tool RUL based on trajectory similarity-based prediction (TSBP) and the differential evolution support vector regression (DE-SVR) algorithm. The most representative features of collected cutting force signal were extracted in time-domain, frequency-domain and time-frequency-domain. After the correlation investigation, the selected features were import into the proposed model for prediction. The integrated prediction model has been proved to be effective in predicting tool RUL. An et al. (2020) proposed a model that combines a CNN and stacked bi-directional and uni-directional LSTM (SBULSTM) networks for cutting tool RUL prediction. CNN extracted signal features for dimensionality reduction, and then SBULSTM trained these features and achieved the purpose of prediction. In order to verify the integrated model, a machining experiment of vibration and current signals have been adopted. Although the model is proved feasible to be used to monitor tool status and predict the RUL, since the model contains many layers and the structure is complex,

there is still a distance for its application in the real environment. These literatures are summarised in Table 2.

Table 2. The literature of cutting tool RUL prediction

Sensor signal	RUL prediction method	Data partition	Reference
Vibration, cutting force, AE	ANN	×	Karam et al. (2016)
Vibration in three-axis	NFN	×	Zhang et al. (2016)
Vibration	HMM	HMM	Yu et al. (2017)
Vibration, current, AE,	ANFIS	Manually segmentation	Wu et al. (2018)
Cutting force	TSBP+ DE-SVR	RMS, AV, standard deviation	Yang et al. (2019)
Vibration, current	CNN+LSTM	Tool wear region threshold	An et al. (2020)

2.4 Sensor signal processing technology

2.4.1 Feature extraction and feature selection

The development of intelligent tool condition prediction models has led to a surge in the amount of data in data mining tasks, and high-dimensional data often harms learning models. The use of effective signal features outperforms the combination of excellent models and inferior features even in poor training models (Khan et al. 2019). Feature extraction and feature selection are widely adopted methods, which aim to obtain valuable features from original feature set according to specific evaluation criteria. Although, the independent employment of these two methods is feasible, their combined application will improve the prediction performance significantly that has been proved in many TCM related studies (Chen et al. 2019, Mohanraj et al. 2019, Plaza et al. 2019).

Since the purpose of feature selection is to obtain a feature set with representative information, and the collected massive data of sensors in TCM system is non-stationary and time-varying that cannot be directly adopted as the feature set. It is necessary to extract the signal features with obvious practical and statistical significance, which are corresponding to the dependent variables, to boost the efficiency and accuracy of tool status recognition. Moreover, in the sensor signal

processing research of the TCM, tool wear leads to the change of amplitude and fluctuation characteristic of the signal, and the statistical features can better reveal the inherent attributes of the data. Therefore, the statistical features have become a popular choice for feature extraction to capture the tool wear status (Feng et al. 2019). Prevalently utilised statistical feature, such as the standard deviation is a measure of the power of a sensor signal and effectively represents the performance of the cutting tool. The kurtosis, which indicates the sharpness or flatness of the signal (Niu et al. 2019), establishes the relationship with the condition of the cutting tool through the change of peak value. Skewness defines the degree of asymmetry around the mean of the distribution, which is affected by cutting conditions and tool wear. For the maximum and minimum value, they reflect the fluctuation of the signal amplitude to predict the tool wear (Madhusudana et al. 2016). The statistical features that have been adopted in recent research based on multi-sensor TCM are summarised as Table 3.

Table 3. The literature of feature extraction

Sensor Signal	**Extracted feature**	**Reference**
Cutting force, AE, Vibration	Mean, RMS, Kurtosis, Variance	Wang et al. (2015)
Cutting force, Vibration	Mean, Maximum, Peak to peak, RMS, Variance, Impulse factor, Kurtosis factor, Margin factor, Shape factor, Crest factor, Skewness factor	Liu et al. (2015)
Cutting force, AE, Vibration	Maximum, Peak to peak, RMS, Skewness, Kurtosis, Variance	Wang et al. (2016)
Cutting force, AE, Vibration	Mean, Standard deviation, Peak to peak, RMS, Skewness, Kurtosis, Impulse factor, Margin factor, Shape factor, Crest factor	Yang et al. (2016)
Cutting force, AE, Vibration	Maximum, Peak to peak, RMS, Skewness, Kurtosis, Variance	Wang et al. (2017)
Cutting force, Vibration, Tool tip temperature	Mean, Standard deviation, RMS, Skewness, Kurtosis, Variance	Rizal et al. (2017)
Cutting force, AE, Vibration	Mean, Skewness, Kurtosis, Variance	Caggiano (2018)
Cutting force, AE, Vibration	Mean, Standard deviation, RMS, Kurtosis Factor, Margin factor, Shape factor, Crest Factor, Skewness Factor	Zhou and Xue (2018)
Torque,Thrust force	Mean, Maximum, Minimum, Standard deviation, RMS, Skewness, Kurtosis, Variance	Duo et al. (2019)
Current, AE, Vibration	Mean, Standard deviation, Peak to peak, RMS, Skewness, Kurtosis, Mean absolute deviation (MAD), Median, Variance, Impulse Factor, Margin Factor, Crest Factor	Lee, et al. (2019)

Furthermore, as an essential part of data fusion technology, feature selection can eliminate redundant or irrelevant features, especially for extensive multi-source high-dimensional data. Without the elimination, the prediction algorithm will incorrectly build the connection between redundant features and response variables, which increases the prediction error. On the other hand, the feature selection reduces the feature dimensionality, and it is conducive to the interpretability and computational feasibility of the learning model. Meanwhile, a limited number of features helps avoid overfitting (Benjamin et al. 2018). The advantages of feature selection can be summarised as follows:

- Reduce the dimensionality of the feature set to ensure data storage demands.
- Remove redundant, irrelevant or noisy data.
- Accelerate the computing time of learning algorithms of data analysis tasks.
- Enhance data quality, and thus improves the performance and prediction precision of the learning model.

In view of the significant superiority of feature selection in boosting the prediction model performance, this technology has been applied in many fields related to data processing. The filter method and wrapper method are two types of feature selection mainly used in the field of TCM, the advantages and disadvantages of these two methods are summarised as Table 4.

Table 4. Pros and cons of filter method and wrapper method

	Advantage	**Disadvantage**
Filter method	• Simple and faster computing	• Dependencies between features are not considered • Low accuracy based on the statistical indicators
Wrapper method	• Evaluate feature subset number with cross validation • The use of classifiers has achieved higher accuracy	• Large amount of computation

The filtering method sorts all the features according to the relationship between the original features and the dependent variables, then selects the representative features through the manually set threshold (Hira and Gillies 2015). This method is independent of classifiers and uses statistical indicators to score and filter each feature, which focuses on data itself and retain low computational complexity. However, the weakness of the filtering method is that, it only

considers individual feature and ignores possible interactions between features. Such consequences may result in the acquisition of many highly relevant features, while high redundant information will make classification and prediction performance worse (Chang et al. 2019). In contrast, the wrapper method adopts an estimator to evaluate the potential subsets with different features sequentially, and ultimately obtain the best feature subset. The performance of the wrapper method is more dependable than the filtering method in that, the wrapper method assesses the feature subset through learning algorithms which their higher accuracy helps to find the optimal feature subset (Mao and Yang 2019). The research of popular feature selection algorithms of the above two methods in the TCM area is summarised as Table 5.

Table 5. The literature of feature selection

Feature selection type	Algorithm	Sensor Signal	Reference
Filter method	PCC	Vibration	Zhang et al. (2016)
		Current	Neef et al. (2018)
	mRMR	Cutting force	Wang, Yang and Guo (2013)
	ANOVA	Vibration, Cutting force	Mali et al. (2017)
	Fisher's discriminant ratio (FDR)	Cutting force, Vibration, AE	Geramifard et al. (2012)
		Sound, Voltage	Zhang et al. (2015)
		Vibration, Cutting force	Chen et al. (2018)
	Entropic measure	Vibration	Painuli et al. (2014)
Wrapper method	Genetic algorithm	Cutting force, Vibration, AE	Zhou and Xue (2018)
		Cutting force, Vibration, AE	Pandiyan et al. (2018)
	Stepwise selection	Vibration	Simon and Deivanathan (2019)
	Decision tree	Vibration	Madhusudana et al. (2019)

It is worth to notice, based on the literature in earlier years that the filter method was used, Zhang et al. (2016) mentioned that the PCC only constructs linear correlation between features

and tool wear because of its uncertain assumptions. In the research of the tool wear status classification, Xie et al. (2019) found that the features selected by FDR only reflect severe wear, and not sensitive to other wear states. Nevertheless, the wrapper method is gradually employed and becoming a more enthusiastic choice for researchers.

Regarding the feature selection algorithms belonging to the wrapper method, recursive feature elimination (RFE) is the most effective one. The RFE algorithm offers significant advantages for nonlinear classification problems because it can perform group prediction analysis on multi-dimensional features and then control the selection of feature subsets (Khaire and Dhanalakshmi 2019). The concept of RFE is to perform repetitive training with a learning classifier. In each round of training, the worst feature will be deleted according to importance ranking, and then the remaining features will be trained for the next round. This recursion continues until all features have been processed (Chatterjee et al. 2019), and the feature that is finally eliminated is considered the most relevant to the response variable.

For RFE feature selection, there are more than one classifier can be selected (Pes 2019). The SVM is a powerful tool for prediction models and classifiers. Its main idea is to mapping samples into higher-dimensional feature spaces, then the decision plane separate elements into different classes. Additionally, SVM can handle a large amount of data and provides excellent performance for nonlinear decision making (Nalepa and Kawulok 2019). In the study of feature selection methods, various classifiers, including SVM, have been verified. Lee et al. (2012) performed the comparison of four feature selection methods for colonic lesions monitoring, in terms of the filter method and wrapper method. The results show that the SVM is superior to the gradient boosting machine, random forest (RF), and linear discriminant analysis. Lai et al. (2017) tested three feature selection methods based on temporal lobe epilepsy magnetic resonance image classification, namely t-test filtering, sparse-constrained dimensionality reduction model, and RFE-SVM. SVM shows significant performance among these three methods. Moreover, Huang et al. (2017) reviewed the progress of SVM-based monitoring in cancer genome research, and they mentioned that SVM stands out from numerous methods in feature selection. Wu and Faisal (2019) proposed a study on the feature selection of breast cancer data based on RFE-SVM, and the proposed method achieved the prediction accuracy of 97.19%. Different from the above literature, SVM is more often used as decision-making model in recent research related to TCM (Sun et al. 2019, Niu et al. 2019, Zhou et al. 2020), not being valued at the feature selection level.

2.4.2 Feature selection with deep learning

Feature selection as an effective data processing method is often incorporated with machine learning models to perform decision-making. Research based on deep learning algorithms usually

skip the feature selection directly (Li et al. 2018, Luo et al. 2019). Without feature selection, the random perturbations of the input will significantly affect the performance of deep learning models (Ghorbani et al. 2018). Recent related studies have shown that feature selection technology promotes the performance of deep learning prediction models.

Lu et al. (2018) proposed a system that integrates feature selection and DNN, this system was applied to real data experiments, in terms of identifying drug resistance, identifying nutrient intake and body mass index, the validation results display that the proposed framework was two times more efficient than the framework without feature selection. Based on the study of ovarian cancer and breast cancer image classification, Chen et al. (2019) evaluated the influence of feature selection technology on the performance of three deep learning models of CNN, DBN (deep belief network) and RNN (recurrent neural network). First, an experimental verified their hypothesis, which the addition of feature selection effectively improves the accuracy of deep learning prediction. Secondly, they compared 11 feature selection algorithms, such as t-test, chi-squared test and RF, etc., among them, REF-SVM achieved the best prediction accuracy.

Not only in biomedicine area, but there is also research in the combination of feature selection and deep learning networks in the area of machinery diagnostics. Maurya et al. (2019) proposed a fusion model of feature selection and DNN, based on three datasets: AE dataset collected in reciprocating air compressors, vibration signal collected from groove ball bearings and the dataset of steel plate fault, they verified three feature selection methods: RF, radial basis function SVM and linear SVM separately. Finally, the prediction results of DNN show that the scheme using feature selection for machine condition prediction has a considerable improvement in performance.

Although numerous research works have focused on the feature selection, the scenario of feature selection in combination with deep learning algorithm in the field of TCM has been paid less attention; research has more focused on the field of biomedicine, relatively. Chen et al. (2019) concluded that, within 195 authoritative biomedical papers that related to deep learning models in recent years, 36 have matched with feature selection. Therefore, similar research is worthwhile in the direction of TCM.

2.4.3 Hurst exponent

In multi-sensor signal based prediction, a reasonable prediction not only depends on an appropriate learning model, but the form of input signals will have a significant impact. Hurst exponent is an index used to estimate the fractal characteristics of time-series signals, and it is capable to effectively segment sensor signals to improve the quality of the data. Hurst exponent

was developed by British hydrologist H. E. Hurst (1900-1978) based on the rescaled range (R/S) analysis method, and the initial purpose was to study the correlation between the flood of the river Nile and the drought process (Luo and Huang 2018). With continuous improvement, the Hurst exponent is extensively applied to determine the long-term memory of time-series data, which is the varying degree of the time-series data within the time span (Knight and Nunes 2018).

Thanks to effectively distinguish the random signals and cross-correlated signals, Hurst exponent has been successfully used in the chaotic fractal analysis of capital markets (Ramos-Requena et al. 2017, Kroha and Skoula 2018), Internet traffic analysis (Dymora and Mazurek 2019), seismic signal analysis (Zhang et al. 2015, Ortiz et al. 2016, Sarlis et al. 2018), pathological index analysis (Jing et al. 2017, Marusina et al. 2017, Dong et al. 2018), etc. Focus on the machinery condition monitoring and prediction field, Vela-Martínez et al. (2009) adopted Hurst exponent to monitor the instability of vibration in cutting tools. By verifying the acceleration signal during the milling process, the results were consistent with the theoretical cutting stability under different machining conditions. Hurst exponent was used as an index of cutting stability in their study, and it has been proven to be effective in confirming the fractal change of the sensor signal, which refers to the large degree self-similarity of the signal. In addition, Sun and Zhao (2011) proposed a monitoring system of cutting tool wear based on cutting surface texture images processed by the Hurst exponent. The Hurst exponent was extracted from the cutting images captured in the cutting experiment. The experiment shows that the Hurst exponent keeps decreasing with the tool wear, which indicates the signal trend is predictable and Hurst exponent shows a good expression ability to judge the tool wear process. The research of Lin and Chen (2014) applied the crossover properties for fault diagnosis, and proposed a nonlinear data feature extraction method to diagnose rotating machinery faults. First, the detrended fluctuation analysis was adopted to assess the vibration signal from a rotating machine. Then, these signal curves were divided into several different scaling regions by the Hurst exponent. Finally, the Hurst exponent values corresponding to the signal segment curves were used as features, to realise the evaluation with the real defective rotating machinery. The vibration datasets of an actual gearbox and rolling bearing were used to validate the effectiveness of the proposed scheme. The result shows that the method performs well in identifying the fault types of rotating machineries. Guan et al. (2018) proposed a method of monitoring tool wear condition based on multifractal de-trended fluctuation analysis and least squares SVM. The proposed method first analysed the fractal spectrum of the AE signal by the Hurst exponent, to characterize the fluctuation of the signal under different wear conditions, and then constructed the clustering feature of the tool wear. Finally, the prediction of tool wear was executed by SVM and reached a satisfying accuracy. They concluded that the Hurst exponent can be used to characterize the characteristics of tool wear by clearly distinguishing tool wear stages. Mohanty et al. (2018) constructed a bearing fault prediction model using vibration

and AE signals. The model first applied variational mode decomposition and empirical mode decompositions (EMD) to decompose the vibration and AE signal in the frequency spectrum, and then adopted the correlation coefficient and Hurst exponent to determine the bearing fault from the decomposed signal. The model has been verified in cases with different speeds, and the proposed method shows great potential in detecting bearing failures. The above literature related to Hurst exponent is concluded in Table 6.

Table 6. The literature of Hurst exponent in machinery aspect

Prediction object	Prediction signal	Reference
Tool wear	Vibration	Vela-Martínez et al. (2009)
Tool wear	Workpiece surface texture image	Sun and Zhao (2011)
Rotating machinery fault	Vibration	Lin and Chen (2014)
Tool wear	AE	Guan et al. (2018)
Bearing fault	Vibration, AE	Mohanty, Gupta and Raju (2018)

According to the relevant literature of the Hurst exponent, it has already displayed its maturity and reliability of data analysis in various fields. However, the utilisation of the Hurst exponent in machinery condition monitoring is still reserved, which is worthwhile to explore rely on its following advantages (Kristoufek 2010, Oral and Unal 2019):

- Hurst exponent is capable of finding the self-similarity and long-term memory of signal sequences with time, without assuming the distribution characteristics of time series.
- Hurst exponent reflects the self-similarity degree of data based on the statistical features. It is easy to execute, can be realised simply with programming, the computing speed is fast and not susceptible to the influence of time scale. Thus, it has good scalability and robustness.
- Hurst exponent has been verified for a long time and more reliable than other methods.

2.4.4 1D-to-2D signal format conversion

Inspired by plenty of successful cases of CNN in the image recognition, many researchers have adopted the data processing strategy of converting 1D time-series data to 2D form in the application of machine state diagnosis based on sensor signals. Li et al. (2017) developed a CNN-based bearing diagnosis method, which constructed the image input with the RMS of vibration frequency-domain data. Tra et al. (2018) adopted the same signal format conversion method based

on AE signals, and obtained notable results in bearing diagnosis. Ding and He (2018) proposed a scheme of employing the WPT (wavelet packet transform) to convert vibration signals into a 2D time-frequency-domain image to realise bearing fault diagnosis. The proposed system performs better in the evaluation compared with other feature extraction methods. However, they only used a single signal source. Chen et al. (2019) proposed a recognition system of planetary gear state, which directly spliced vibration signals from different monitoring positions into a 2D matrix to perform diagnosis on a CNN model. The verification results display that the classification accuracy of this system is improved compared to the unconverted 1D signal. Nevertheless, the diagnostic information provided by a single sensor signal in the research still insufficient. Similarly, Yang et al. (2019) constructed a fault diagnosis system for reciprocating compressors, and three-axis vibration signals were merged into a 2D matrix as the input of a CNN model. Wang et al. (2019) proposed a method for diagnosing rolling bearings based on vibration signals. The collected signals were converted into a pixel matrix to generate a 2D grey image. The final prediction accuracy of CNN model reached nearly 100%. This robust and concise method was also used in the diagnosis of wind power system and centrifugal pump. Cao et al. (2019) proposed a tool status recognition system. Before the CNN-based diagnosis, the only input signal of the system, spindle vibration signal, was converted into frequency-domain through derived wavelet frames, and then these data were stacked into 2D form. This method provides higher recognition accuracy than the use of raw data.

2.4.5 Data fusion in deep learning algorithms

Deep learning is emerging as a vital technology of IoT application systems. It shows better performance at the scale of large data amount and makes EC more intelligent and efficient. Moreover, the use of data fusion will further maximise the accuracy of deep learning when processing multi-sensor data from a wireless sensor network (WSN). In addition to the area of smart city (Meng et al. 2017, Cao et al. 2020), the radar recognition (Jia et al. 2019, Zhou et al. 2019), medical image (Li et al. 2020) and so on, the system architecture that incorporates deep learning and data fusion has also played an advantage in the field of industrial monitoring and diagnosis.

Zhang et al. (2017) presented a condition monitoring system of ball screw relied on DBN and multi-sensor data fusion. First, raw data from multiple accelerometers were converted to frequency-domain, and the frequency spectrum of the signals was fused with a parallel superposition method. Then, a DBN learning model was built according to the fused frequency-domain features. The dataset of a real ball screw degradation experiment was used to conduct

comparative assessment between the fused and unfused signals. The result displays that the proposed system provides satisfied recognition accuracy.

Chen and Jahanshahi (2018) proposed a CNN-based deep learning framework to monitor cracks in the metal surface of nuclear power plant reactors, in which Naivebayes was employed to perform data fusion on each video frame that captured per second, and then CNN was developed to use this fusion information to detect cracks. Collected actual video samples was adopted to evaluate the effectiveness of the proposed system, this deep learning model achieves an accuracy of 98.3%, and the response is rapid compared to the SVM-based method.

Yao et al. (2018) presented a deep learning method of gear failure identification. The model first extracted the features of acoustic signals under time-domain and frequency-domain, and these features were merged through matrix combination, then the gear failure was detected through the established CNN model. The experimental results of the two datasets show that the identification accuracy of the proposed model reaches 98.5% and 96.5%, respectively, and its performance is higher than traditional fault diagnosis methods.

Chen et al. (2019) proposed a health monitoring model for the two-stage planetary gearbox. In this method, the raw vibration signals of horizontal axis and vertical axis were combined, and then features of combined data were automatically extracted by CNN to complete the final fault diagnosis. The CNN-based method was compared with SVM and back-propagation neural networks with a real dataset. The result presents that the classification precision of the proposed model reaches 90%, which is much higher than other methods.

Shi et al. (2019) proposed a deep learning framework of tool condition monitoring for the ultra-precision machining tool. The model applied multiple sparse autoencoders to achieve feature fusion, and to build the relationship between the features of sensor signals and tool status. The vibration signals from experiments in real factories were adopted to validate the proposed model. The result reveals that the method integrated data fusion is more accurate and robust than the standard deep learning method, and the classification accuracy of ultra-precision machining cutting tool is more than 96%. However, this method has certain limitations for different tool objects.

Liu et al. (2019) proposed an ensemble 1D CNN bearing fault diagnosis model. First, the raw data from sensors was converted to frequency-domain, and the frequency spectrum of the signal was then input into sub-CNN model for false identification. Finally, the SVM merged the result of the multi-CNN models and output the integrated results. The datasets of two real bearing test

experiments were used to verify the performance of the model. The result shows that the proposed method provides higher recognition accuracy and avoids the information discarding.

He et al. (2020) presented a rotating machine condition diagnosis strategy relied on vibration signals and CNN. The vibration signals from each acquisition point were input into multi-channel CNN models to perform the identification of different faulty types, and then the decision was fused by the majority voting for the complete result. The validation of this method was based on a real dataset, and the result shows that this method is advantageous in faulty status recognition.

Driven by big data, Zhu, Li and Zhang (2020) proposed a system of intelligent TCM. In this framework, the signal features of various sensors were extracted, and then the generated features were combined as a new feature set to achieve the fusion. Finally, the deep learning prediction model composed of RNN and CNN realised the tool life prediction based on the feature set. The system was verified through a case experiment dataset of cutting force, vibration, AE signals and tool wear images, the result shows that the proposed multi-signal fusion and deep learning method significantly improve the tool wear recognition accuracy and the performance of intelligent TCM system.

Wu et al. (2020) put forward a RUL prediction method relied on deep LSTM and multiple sensor signals. The model simultaneously combines the ability of sensor signal fusion and prediction. Two turbine engine datasets were used to verify the presented method, and the result shows that the method adaptively estimate the engine RUL based on the acquired data of 100 sensors, the prediction accuracy is satisfactory. The above works are summarised in Table 7.

Table 7. The summary of relevant studies in data fusion and deep learning

Fusion method	Deep learning algorithm	Collected Signal	Prediction object	Reference
Parallel superposition	DBN	Vibration	Ball screw condition	Zhang et al. (2017)
Naivebayes	CNN	Image	Cracks in the metal surface	Chen and Jahanshahi (2018)
Matrix combination	CNN	Acoustic	Gear failure	Yao et al. (2018)
CNN	CNN	Vibration	Two-stage planetary gearbox	Chen et al. (2019)
Sparse autoencoders	Multiple sparse autoencoders	Vibration	Tool condition	Shi et al. (2019)
SVM	CNN	Vibration, current	Bearing fault	Liu et al. (2019)
Majority voting	CNN	Vibration	Rotating machine condition	He et al. (2020)
Data mergence	RNN + CNN	Force, vibration, AE	Tool condition	Zhu, Li and Zhang (2020)
LSTM	LSTM	100 sensors	Turbine engine RUL	Wu et al. (2020)

2.5 Deep learning algorithms

2.5.1 CNN

As a deep network, CNN is widely recognized for its convolution layers. Besides, CNN also involves pooling layers, fully connected layers and flatten layers. CNN was first proposed by LeCun et al. (1998) for image processing, and it has shown its ability to solve complex cases in various applications, such as computer vision (Fang et al. 2019, Xu et al. 2020) and speech processing (Amin et al. 2019, Fujimura et al. 2020), since it can maximise the retention of raw data information. Benefiting from the shared weights and pooling functions of CNN, high dimensional signal can be directly used to obtain excellent training result within an acceptable time cost (Yamashita et al. 2018). Over the years, the CNN model has been effectively employed in machinery fault monitoring and diagnosis.

A bearing fault state monitoring model based on CNN was proposed by Janssens et al. (2016). Taking the vibration signal as input date, they compared the performance of the feature selection process on a CNN model, and it shows that the classification accuracy with performing feature selection is improved. However, the use of only one sensor signal increase the possibility of misclassification. Liu et al. (2017) presented an intelligent motor fault diagnosis system based on CNN. This system used the dislocated time-series CNN to shift the input original signals to extract signal features continuously. By applying the vibration signals as an indicator for the validation, the result shows that the diagnostic framework could provide a stable prediction for motor fault. To realise the fault diagnosis of aero-engine sensors, Li and Qu (2018) put forward a CNN diagnosis model. In this model, PCA reduced the dimensionality of the simulated sensor signal, and then a CNN model was constructed to execute the prediction. Through the comparative analysis of simulation experiments, it implies that this method can achieve satisfactory accuracy in aero-engine sensor fault diagnosis. Huang et al. (2019) proposed a CNN-based multi-domain prediction method of tool wear. The features of cutting force and vibration signals in time-domain, frequency-domain and time-frequency-domain were extracted as indicators, to establish the relationship with tool wear on the CNN model. The experimental result presents that the prediction precision of the proposed model is significantly higher than other methods. However, the signal features in this method are still manually extracted based on prior knowledge, the real-time performance of the model is limited. Martinez-Arellano et al. (2019) applied the GASF to encoding original multi-sensor signals as images, and attached the images to a CNN model to realise tool wear classification and achieved a considerable classification accuracy. This method avoids the manual extraction of signal statistical features, ensures the integrity of data information, and more suitable for the processing of big data. However, the complexity and professional knowledge demand become the limitation of this model for practical use. The summary of the above literature is shown in Table 8.

Table 8. The literature of CNN in machinery prediction aspect

Prediction object	Prediction signal	Reference
Bearing fault	Vibration	Janssens et al. (2016)
Motor fault	Vibration	Liu et al. (2017)
Aero-engine sensors	Simulated sensor	Li and Qu (2018)
Tool wear	Cutting force, vibration	Huang et al. (2019)
Tool wear	Cutting force, vibration, AE	Martinez-Arellano, Terrazas and Ratchev (2019)

2.5.2 CNN architecture

According to the different dimensionality of the input variables, CNN is commonly classified into 1D and 2D structures, which the 1D CNN principally used for time-series processing, and 2D CNN is more suitable for image perception (Yamashita et al. 2018). In the field of machinery condition prognosis, to process the machinery time-series signal with 1D CNN is appropriate, it better utilises the characteristic of CNN in automatic feature extraction (Kiranyaz et al. 2019). Thanks to the potential of all aspects of 1D CNN in processing sensor signals, it has appeared in many research in the field of machinery monitoring based on multi-sensor signals (e.g. Jing et al. 2017, Zhao et al. 2019 and Li et al. 2020). 1D CNN has advantages in the following aspects (Kiranyaz et al. 2019):

- 1D CNN is array operation instead of matrix operation. Thus, the configuration is concise and compact, and the complexity of computation is low.
- The shallow-architecture of 1D CNN network contains a small number of neurons and layers, which is easier to train and implement.
- 1D CNN is achievable on standard computer CPU (central processing unit), it does not require high performance and particular hardware configuration.
- 1D CNN is suitable for real-time and low-cost applications, because of the low computational demand.

On the other hand, according to specific needs, 2D CNN is also applied to process sensor signal and brings positive effects on the accuracy improvement of the prediction. For instance, in the work of (Janssens et al. 2016, Sun et al. 2017 and Martínez-Arellano et al. 2019), they converted sensor signal into image format for matching the use of 2D CNN, which is a necessary operation for 2D CNN to process time-series data (Fawaz et al. 2019). However, the similarity drawback of these studies is that they adopted the format conversion scheme only on a single sensor signal.

2.5.3 LSTM

Among the field of deep learning, in addition to feed-forward neural nets like CNN that perform well on the continue nonlinear data, the feedback neural network, which redistributes weights by feeding back information from the output layer to each layer to reduce the final prediction error (Herzog et al. 2020), is proposed to analyse the logic sequence between the input data with different lengths and time correlation. As the most representative feed-back neural, LSTM has become a promising method to handle the long-term signal, which was first proposed by Hochreiter and Schmidhuber (1997). It should be mentioned that LSTM is a unique structure

of RNN. The RNN establishes a connection between input samples. However, the problem of gradient disappearance during model training limits the performance of RNN, that not capable of maintaining long-term dependence within sensor data. Therefore, the forget gate has been introduced in LSTM to improve the problem of RNN, the forget gate decide whether to discard non-essential information and reduce the negative impact on subsequent predictions. Considering that LSTM can directly capture the dependence and nonlinear relationship in time-series data without the need of additional manual intervention, it has been applied and achieved success in some areas including image recognition, text recognition and activity recognition (Mikolov and Zweig 2013, Stollenga et al. 2015, Qin et al. 2017, Xu et al. 2017). At the same time, due to the inherent time sequence properties of sensor signals, LSTM has also been widely used in the machine monitoring area.

Zhao et al. (2016) studied the performance of LSTM-based machine health monitoring system. In the study, the cutting force, vibration and AE signals were adopted to map cutting tool status, and the performance of single-layer and multi-layer LSTMs were verified, respectively. The experimental result proves that LSTM is superior in machine health monitoring. In order to monitor the mechanical state of motor shafts, Chen et al. (2017) proposed an LSTM prediction model of motor bearing state. The stable values of vibration signals were obtained through EMD and used for prediction. Compared with the results of SVR, LSTM achieves higher accuracy of the machine state prediction. Wang et al. (2018) constructed a fault diagnosis framework based on LSTM. The research took gear failure as the experimental object, and adopted the vibration signals as input for prediction. The performance of the proposed LSTM method has better classification effect than SVM and RNN, the prediction accuracy increases to 99.80%. Lei et al. (2019) proposed a state monitoring model of wind turbines based on LSTM, that used the time-series signals of displacement and vibration signals for fault diagnosis. Due to the adoption of LSTM, this method free from redundant signal processing. Finally, the verification of single-sensor and multi-sensor data shows that the proposed method has obvious advantages. Elsheikh et al. (2019) presented a bidirectional LSTM structure for estimating RUL through observed sensor signals. By taking 26 time-series sensor signals of turbofan engine as input, the result of LSTM prediction displays that the proposed method is effective for RUL prediction based on random starting state. Sun et al. (2020) proposed a prediction model of cutting tool state relied on deep learning model for achieving timely tool replacement. The LSTM network predicted the value of flank wear based on historical data of cutting force, vibration and AE signals. The experimental research shows that the proposed method reliably predicted the cutting tool wear. The summary of the above research is given in Table 9.

Table 9. The literature of LSTM in machinery prediction aspect

Prediction object	Prediction signal	Reference
Tool wear	Cutting force, vibration, AE	Zhao et al. (2016)
Motor bearing	Vibration	Chen et al. (2017)
Gear failure	Vibration	Wang et al. (2018)
Wind turbine	Displacement, vibration	Lei et al. (2019)
Turbofan engine	26 time-series sensor signals	Elsheikh et al. (2019)
Cutting tool state	Cutting force, vibration, AE	Sun et al. (2020)

2.5.4 Hybrid CNN-LSTM

Contrasted with an individual CNN or LSTM model, the application of hybrid CNN-LSTM model is in the early stage. A small amount of existing relevant research mainly focusses on language, image recognition and video processing, etc. For instance, speech recognition (Sainath et al. 2015), action video recognition (Ullah et al. 2017, Zhu et al. 2020), traffic forecasting (Bogaerts et al. 2020), and residential energy consumption (Kim and Cho 2019).

In the mechanical engineering area, as the result of the desire of integrating different deep learning algorithms to solve higher-level diagnosis, the combination utilisation of CNN and LSTM gradually entered the research scope. Gao et al. (2018) proposed a scheme for icing faults prediction of wind turbine blades based on CNN-LSTM model. The collected generator operating data was directly used to diagnose the icing state of the blades after a simple format transformation. The evaluation results display that the proposed CNN-LSTM model has achieved a better classification effect than other commonly used classification models, SVM, CNN and so on. He et al. (2019) proposed a combined model of CNN and LSTM for detecting the fault of rod pumping system. The model took the indicator images of the system as input, and relied on the advantages of both CNN and LSTM model to extract image features to achieve the fault category classification. Finally, by employing the dataset from an oil well, the performance of the proposed model and traditional machine learning prediction models (RF, linear regression, etc.) were evaluated, the results show that using the CNN-LSTM model to perform fault classification is more effective. In order to realise the prediction of turbofan engine RUL, Kong et al. (2019) proposed a scheme based on a hybrid CNN-LSTM model. First, multivariate data from the engine was processed through a polynomial regression method, to obtain a health index of engine

degradation. Then, the index and dataset were fed into the established model, which performs spatial feature extraction and time-series feature extraction in sequence, and finally achieve the regression prediction. The effectiveness of the introduced model was verified using simulation data from NASA Commercial Modular Aero-Propulsion System. In addition, in the performance comparison with MLP (multilayer perceptron), SVR, CNN and LSTM models, the hybrid CNN-LSTM model performs better. These literatures are summarised in Table 10.

Table 10. The literature of CNN-LSTM in machinery prediction aspect

Prediction object	Prediction signal	Reference
Wind turbine blade icing faults	Generator operating data	Gao et al. (2018)
Rod pumping system fault	Indicator image	He et al. (2019)
Turbofan engine RUL	Multivariate sensor data and parameters	Kong et al. (2019)

2.6 IoT network component

2.6.1 Economic TCM

Limited commercial TCM systems are useable, nevertheless, they are not popular in SME's due to the high expense. Thus, the study of low-cost multi-sensor monitoring system appears in some fields in recent years, such as environment monitoring (Khoa and Takayama 2018, Bamodu et al. 2018, Intrieri et al. 2018), gear vibration (Wu et al. 2012), water quality detection (Alahi et al. 2018, Lambrou et al. 2014), human health monitoring (Garbhapu and Gopalan 2017, Harbouche et al. 2017) and so on. Although, the research of the development of economical machinery status monitoring system is rare in the field of tool conditions, it has not stopped.

Lee (2006) built up a neural network-based TCM system that to decrease the breakdown of turning machine caused by tool wear, the cost-efficient vibration and AE sensors have been utilised in the system, and the general accuracy of tool condition prediction was implemented.

Ghani et al. (2011) developed a low-cost tool wear monitoring system. The system collected cutting force signal by strain gauge sensors to detect and analyse the deflection of cutting tool corresponding to the tool wear. MATLAB was used as an online monitoring user interface. In the experiment, the Z-filter technique was applied to construct the relationship between the sensor signals and flank wear, then to achieve the prediction. The results indicate that the proposed monitoring system is effective and affordable to detect flank wear during machining.

Huang et al. (2016) developed a low-cost platform for measuring machine vibration using the ADXL001 MEMS accelerometer, they evaluated the effectiveness of the platform according to the ISO-16063-21 calibration method. Finally, a milling experiment was performed to verify the performance of the developed vibration monitoring platform and a platform using the PZT commercial accelerometer. By contrast, the proposed strategy is more suitable for machine tool state monitoring.

Mail et al. (2017) developed a monitoring system of tool wear relied on the force and vibration sensors, they have validated the multi-sensor platform with a Taguchi experiment, the result shows that the vibration and force signals have a consistent correlation with the tool wear. They also emphasised that the multi-sensing method will ensure higher accuracy of tool wear monitoring.

For the tool wear recognition, García-Ordás et al. (2018) presented an online portable and low-cost vision system based on machine learning. The system obtained local texture features from multiple cutting tool edge regions to determine the degree of tool wear. The dataset of the edge profile images of the cutting tool has been used for validation, the result shows that the system is promising in automatic tool wear monitoring.

Villalonga et al. (2019) proposed a visual framework for machine tool state monitoring, aimed at the management and diagnosis at factory level. Based on CC, this framework analysed variables such as temperature and time obtained from PLC (programmable logic controllers), CNC, and robot arms, achieved global status monitoring and decision-making. Moreover, the developed framework supports the use of low-cost platforms.

To prevent the costly manufacturing of outdated CNC machine tools and to move toward Industry 4.0. Hesser and Markert (2019) proposed a programmable sensor monitoring platform to monitor the tool wear of CNC milling machines. In addition to adopting low-cost and all-in-one sensor technology, the platform can exchange information via cloud servers. Furthermore, ANN was used to classify the tool wear. This study used an accelerometer to collect vibration signals from a milling experiment. The experiment proved that with the support of computational intelligence and big data, it is feasible to embed the old machine tools with a programmable sensor monitoring platform to integrate them into Industry 4.0. The above literature is summarised in Table 11.

Table 11. The literature of low-cost machinery condition monitoring platform

Monitoring object	Sensor signal	Method	Reference
Tool wear	Vibration and AE	Neural network	Lee (2006)
Tool wear	Cutting forces	Z-filter	Ghani et al. (2011)
Machine tool	Vibration	ISO-16063-21 calibration method	Huang et al. (2016)
Tool wear	Vibration, cutting force	Taguchi	Mail et al. (2017)
Tool wear	Texture image	SVM	García-Ordás et al. (2018)
Machine tool	Temperature and time	CC	Villalonga et al. (2019)
Tool wear	Vibration	ANN	Hesser and Markert (2019)

2.6.2 Edge computing

The research of the National Cable and Telecommunications Association (NCTA) forecasted that there will be approximately 50.1 billion IoT devices connected to the Internet in the world by 2020 (MohanJussi and Kangasharju 2017), it also poses challenges to the data analysis and security for centralised cloud servers. EC is therefore proposed to implement a distributed computing model, that migrates the sensor data processing from cloud sever to edge computing end, which performs local processing close to the data source without massive data uploading (Rehman et al. 2017), thereby, eliminating high network latency, improving the transmission efficiency and information security level. In the framework of traditional CC centre, cloud server can provide computing and storage services, however, EC is needed to enhance the IoT capabilities for large amounts of data. Although, their service goals are similar, EC is an extension of CC, not a substitute. Table 12 shows the difference between CC and EC (Dhingra et al. 2020, Ning et al. 2020).

Table 12. Comparison of cloud and edge computing

Attribute	Cloud computing	Edge computing
Computing type	Centralized	Distributed
Deployment	High cost	Low-cost
Task type	Large-scale computing or data storage	Real-time processing
Data position	Computing centre only	Computing centre and EC devices
Target user	Internet user	Internet user and local user
Information safety	Low	High

Compared with CC, EC combines the collaborative work of CC and EC to obtain a mutually serviceable computing framework, and it better supports the application of IoT. Its superiorities are highlighted in the following aspects (Tseng et al. 2018, Fan et al. 2018):

- Accelerate the processing and transmission of sensor data, alleviate the network bandwidth pressure, and increase the response speed of user requests for services with lower network latency.
- For manufacturing workshops or enterprises, the cost of configuring local EC is much less than building a more complex CC centre.
- Even if the data terminal is temporarily disconnected from the CC centre, the entire system will maintain communication because of EC, and it ensures the robustness of the system.
- The protection of information privacy is much improved. EC facilitates the client to conduct data differential treatment.

As a promising technology, EC is actively being adopted in various applications to mitigate intensive computing and meet broadband needs. Such as smart firefighting (Wu et al. 2017), military domain (Ehala et al. 2017), driving behaviour monitoring (Azar et al. 2019), Epileptic EEG(electroencephalogram) (Hosseini et al. 2020), smart agriculture (Grady et al. 2019, Bhargava et al. 2019) and vehicle auto navigation (Iqbal et al. 2018). Moreover, to address the low efficiency of conventional IoT in big data processing and transmission, EC prior performs data processing on the front end of the cloud server, to maximise the target state information. Various data processing methods have been employed on the EC end recently. Satija et al. (2017) introduced a filtering method for ECG (electrocardiogram) signals, which is based on discrete

Fourier transform (DFT) and signal turning point (TP). This method was implemented on the EC end that built with Android phone, and the system was verified by a real dataset. Pham et al. (2018) introduced a cloud-based home healthcare monitoring system. The system collected a variety of signals of wearable physiological sensor and motion sensor to realise the resident daily activities and health monitoring. Sensor data was first pre-processed through low-pass filtering and statistical feature extraction in an intelligent edge gateway, and then was sent to a private cloud for remote real-time assessment. Lai et al. (2019) proposed an affordable real-time prediction system of air quality, which is relied on the IoT and EC. Raspberry Pi was used as the edge device, it filtered the signal of six air pollutants (such as SO2, NO2) through the Kalman filter algorithm and to conduct short-term prediction. Then, the prediction results were uploaded to the cloud centre for storage and information feedback. The system has been verified with real air quality monitoring datasets, the results prove that EC improves the prediction accuracy and avoids the problem of transmission delay.

In addition to being used more in the above fields, EC has gradually appeared at the factory level in the manufacturing filed. Wan et al. (2018) proposed an intelligent factory energy-sensing model incorporated with EC to realise the scheduling of multi-task equipments. The particle swarm optimisation (PSO) algorithm was directly applied on the Raspberry Pi EC terminal, to establish the relationship between energy consumption and workloads. Although, the verified model can provide quick response to the production line, the scheduling result is not satisfactory since the lack of data processing. Li et al. (2018) presented a deep learning-based identification system for defective products on assembly lines. The system shared the shallow layer of CNN to the EC end, and then the compressed image data captured by the EC end was delivered to the deep layer CNN at the cloud end for final analysis. Simulation experiment proves that the proposed system is efficient and superior to some existing methods. However, the insufficient versatility of this method is a shortcoming. Liu et al. (2019) presented a strategy of EC to monitor the machine state. This method focuses on vibration signals, the WPT and two-dimensional discrete cosine transform (2D-DCT) were employed for signal filtering, and the perceptual hashing technology (PHT) was used to extract the signal information at the edge. The scheme was validated with two bearing wear datasets, their diagnosis accuracy has been improved. Zhao et al. (2020) proposed a system that adopted the IoT and EC technology to monitor manufacturing resources in workshop. The edge gateway in the workshop was based on Raspberry Pi, and it filtered the passively received manufacturing resource signals by the Kalman filter. Then, the supervised learning of genetic tracking (SLGT) method in the cloud server was applied based on the filtered signals to predict the manufacturing resources and estimate its location. The feasibility of the system was validated in a real air-conditioning manufacturing workshop. The above relevant research is given in Table 13.

Table 13. The literature of edge computing application

Research object	Measured information	Edge processing	Cloud server	EC device	Reference
Human heart health monitoring	ECG	DFT, TP	Not mentioned	Android phone	Satija et al. (2017)
Health care monitoring	Wearable sensor	Low pass filter Feature extraction	Remote assessment	Not mentioned	Pham et al. (2018)
Air quality monitoring	Six air pollutants	Kalman filter	Provide feedback	Raspberry Pi	Lai et al. (2019)
Factory equipment scheduling	Energy consumption	PSO	Not mentioned	Raspberry Pi	Wan et al. (2018)
Defective product monitoring	Image	Shallow CNN	Deep CNN	Not mentioned	Li et al. (2018)
Bearing wear	Vibration	WPT, 2D-DCT, PHT	Not mentioned	Not mentioned	Liu et al. (2019)
Machining resource monitoring	Processing Resource signal	Kalman filter	SLGT	Raspberry Pi	Zhao et al. (2020)

Besides the literature summarised above, some studies directly integrate the collected data through edge gateway without performing any data processing methods (e.g. Wu et al. 2017, Chen et al. 2018 and Verderame et al. 2019), or executing simple manual data pre-processing at the EC end to fill missing values and remove redundant values, etc. (e.g. Cao et al. 2017 and Greco et al. 2019). Moreover, some research employed complex approaches such as machine learning on the EC end for data pre-processing of the IoT system but does not consider the feasibility of device performance in actual applications (e.g. Queralta et al. 2019 and Uddin et al. 2019). Obviously, these methods do not apply to the proposed system in this work.

2.7 Summary

This chapter systematically reviews the technology of TCM, and closely focuses on the research direction of this work. Based on the relevant research, there are still shortcomings and challenges to be improved:

- Although some TCM systems achieve the prediction of tool status in different extent with several feature fusion and prediction algorithms, in consideration of the feature selection aspect, the optimal size of the feature subsets is not earned widespread attention. On the other hand, most systems are developed in the specific-designed processing scenario, which not provide comprehensive and scalable strategies, they are difficult to be applied in actual industries.
- Currently, more feature selection and fusion technologies are applied in collaboration with the machine learning model, ignoring the exploration of deep learning algorithms. In addition, the advantages of multi-sensor signals for TCM are self-evident, the research of the feature selection based on multi-source sensor signals is worth expanding.
- There are some studies realised and implemented signal segmentation strategies, while, the promotion and reliability of these approaches have not been verified in tool RUL prediction.
- The existing RUL prediction systems still have apparent defects. For example, there is great uncertainty in the sensor signal features to represent the cutting tool RUL. On the other hand, with the utilisation of industrial IoT, a huge volume of data will be generated during processing, the model based on machine learning algorithms not capable of processing high-dimensional input data.
- Deep learning has been praised on diagnosis and prediction in various fields, but it has not been widely used in cutting tool RUL prediction. In addition, since the sensor signals from TCM systems are time-series, it is necessary to simultaneously extract the space features and time-sequence features via CNN and LSTM model.
- To date, many types of sensor and signal processing technologies have been adopted in the field of TCM to establish economic platforms. However, most of the studies employ single sensor or visual monitoring devices, which is challenging to provide sufficient prognostic accuracy, and only a few studies have focused on the WSN of TCM. It reveals that a more robust and reliable multi-sensor monitoring platform for IoT system is in demand.
- The growing popularity of EC crosses various fields, it demonstrates that the intervention of EC is feasible to prompt the existing IoT system into more intelligent paradigm. However, in the field of tool status diagnosis, there is a small amount of research related to the EC-enabled system, it is in its infancy.
- Simple processing methods and algorithms on EC end are inadequate in the face of the large amount of data generated by machine tools, it undoubtedly limits the application in actual workshop.
- The 2D format conversion of sensor signals and the data fusion deep learning scheme both show advantages in enhancing the precision and dependability of machinery

condition monitoring system. However, there are still some deficiencies in the existing studies. Such as, the adopted signal 2D conversion method neglects to preserve the full-scale information, and some of them are computationally intensive. Moreover, their applications in multi-sensor scenarios and TCM are rare. These circumstances stimulate the research motivation of exploring the signal format conversion and investigating the multi-sensor signal fusion enabled deep learning strategy in EC-IoT monitoring system.

Chapter 3. Multi-sensor Feature Fusion Tool Wear Identification

3.1 Introduction

From the machining mechanism perspective, the cutting tool wear is positively affected by various factors such as materials of cutting tools and workpieces, cutting forces, cutting heats and friction between cutting tool edges with workpiece surface, etc. (Vazquez et al. 2019). The multi-sensor TCM has been widely used to characterize tool wear status, which benefits from these sensors are non-susceptible to the site environment (e.g. cutting fluid), its measurement does not interrupt the processing, easy installation and on-line monitoring feasibility (Bhuiyan and Choudhury 2014). To improve the prediction of low efficiency and low precision but high calculation occupancy rate, which caused by a large amount of data input, this chapter proposes a multi-sensor tool wear identification system based on a novel multi-layer feature fusion technology and 1D CNN prediction model. In this system, the features of the pre-processed sensor signals are extracted from time and frequency-domain, then, the RFECV-SVM (recursive feature elimination and cross-validation) sequentially executes the assessment of the high representative feature subset size and the importance sorting of the features, which corresponds to response variables, to obtain a feature subset with most effectiveness and minimum dimension. After that, the dimensionality of the feature subset is further reduced by the fusion of the Isomap. Finally, the optimal component features are fed into the 1D CNN model for the prediction performance evaluation.

The remaining parts of the chapter are arranged as follows: Section 3.2 introduces the system design, and the methodology of the proposed system, in terms of signal pre-processing, feature extraction, feature selection and dimensionality reduction, meanwhile, the system implementation based on the real dataset are also described in this part. The prediction model development and the validation results of the system are present in section 3.3. The conclusion is in the final section.

3.2 System methodology and implementation

3.2.1 Overall framework of the methodology

In this work, the methodology for predicting tool wear conditions is depicted in Figure 1. It consists of the following steps:

- Signal pre-processing: Signals from multiple sensors are pre-processed to eliminate noises from the signals for quality improvement;
- Feature extraction: Features under the time and frequency-domains will be extracted from the pre-processed signals;
- Selection of optimal features: Extracted features are sent to an RFECV (recursive feature elimination and cross-validation) process to identify optimal features and remove less important (relevant) features for predicting tool wear. Meanwhile, to further enhance the reliability and efficiency of the prediction, the selected features are fused by an Isomap-based method to reduce the dimensionality;
- The optimal features are taken into a CNN model to accomplish tool wear identification.

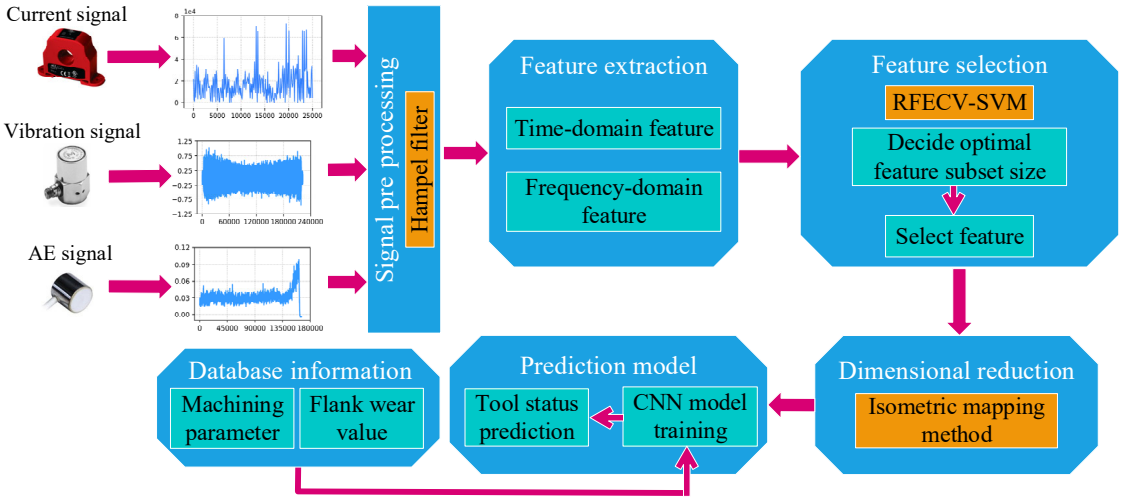


Figure 1. The framework of the proposed methodology

To better explain the methodology, a multi-sensor dataset for milling cutting tools (Agogino and Goebel 2007) are used as a case study. In the case study, three types of sensors, i.e., vibration sensor (the ENDEVCO model 720150), AE sensor (the WD model 925), and current sensor (the CTA model 213), were used to monitor cutting tool conditions. Figure 2 shows the experimental setup sketch, sensors were mounted on the Matsuura machining centre MC-510V. For the setup, two current sensors were used to monitor the AC (alternating current) and DC (direct current) signals of the machine. Two vibration sensors and two AE sensors were mounted on the workpiece table and the spindle of the machine, respectively. Moreover, a 70mm face milling tool, KC710, was adopted during the experiment.

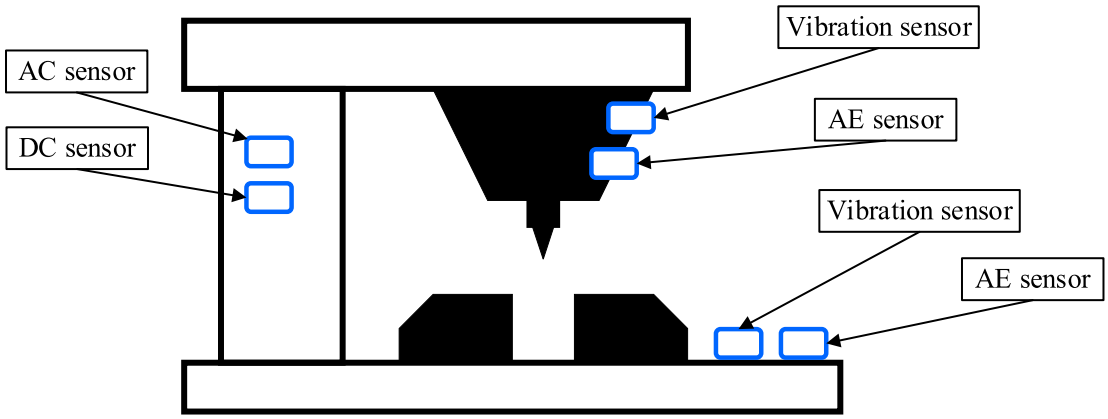


Figure 2. Sensor mounting.

Fifteen run-to-fail cutting cases were conducted in this experiment. The cutting parameters for the experiments, cutting speed, depth of cut, feed rate and material of the workpiece, are listed in Table 14. Each case contains a series of cutting runs, and a total of 164 cutting runs were recorded in the dataset.

Table 14. Machining parameters in the experiment

Case	Cutting speed (m/min)	Depth of cut (mm)	Feed rate (mm/min)	Workpiece material
1	200	1.5	0.5	Cast iron
2	200	0.75	0.5	Cast iron
3	200	0.75	0.25	Cast iron
4	200	1.5	0.25	Cast iron
5	200	1.5	0.5	steel
6	200	0.75	0.25	steel
7	200	0.75	0.5	steel
8	200	1.5	0.5	Cast iron
9	200	1.5	0.25	Cast iron
10	200	0.75	0.25	Cast iron
11	200	0.75	0.5	Cast iron
12	200	0.75	0.25	steel
13	200	0.75	0.5	steel
14	200	1.5	0.25	steel
15	200	1.5	0.5	steel

After each run, the flank wear (Figure 3) of the cutting tool, which occurs at the surface of the tool flank and caused by the abrasion, was measured by a microscope. The measured flank wear of the adopted experimental dataset in this work is shown in Table 15.

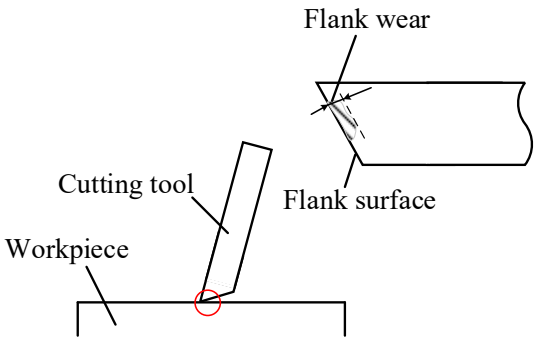


Figure 3. Schematic of flank wear

Table 15. Measured flank wears for experiments

Run No.	Case No.														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	Flank wear V(mm)														
1	0	0.08	0	0.08	0	0	0	0	0	0	0.05	0	0	0.08	0.05
2	0.04	0.14	0.13	0.13	0.16	0.09	0.18	0.1	0.04	0.04	0.08	0.05	0.09	0.15	0.13
3	0.07	0.14	0.13	0.2	0.29	0.13	0.3	0.14	0.08	0.07	0.1	0.1	0.17	0.28	0.24
4	0.11	0.14	0.17	0.31	0.44	0.22	0.36	0.19	0.16	0.07	0.12	0.13	0.24	0.37	0.31
5	0.16	0.15	0.19	0.35	0.53	0.24	0.44	0.27	0.25	0.08	0.17	0.17	0.3	0.48	0.4
6	0.2	0.16	0.2	0.4		0.34	0.62	0.38	0.36	0.09	0.2	0.32	0.35	0.56	0.62
7	0.24	0.18	0.23	0.49		0.46		0.47	0.43	0.1	0.24	0.38	0.6	0.7	
8	0.29	0.22	0.23			0.53		0.64	0.47	0.12	0.32	0.49	0.81		
9	0.28	0.26	0.26					0.81	0.53	0.16	0.36	0.56	1.14		
10	0.32	0.31	0.28						0.7	0.18	0.4	0.68			
11	0.38	0.38	0.33							0.2	0.45	0.83			
12	0.4	0.43	0.36							0.23	0.49	0.92			
13	0.43	0.48	0.44							0.26	0.58	1.07			
14	0.45	0.55	0.55							0.29	0.65	1.3			
15	0.5									0.31		1.53			
16	0.53									0.37					
17	0.54									0.4					
18										0.42					
19										0.47					
20										0.57					
21										0.65					
22										0.68					
23										0.76					

According to the ISO3685:1993, in a metal cutting, the standard threshold for a uniform tool flank wear is 0.4mm, and the threshold for an irregular tool flank wear is 0.6mm. In the subsequent analysis of this work, the values of the flank wear are divided into the unworn set and worn set based on the flank wear threshold 0.4.

3.2.2 Signal pre-processing

During a monitoring process, sensor signals contain random spikes or outlier signals that affect the accuracy of the tool wear identification. The Hampel filter, which demonstrates good performance of removing spikes or outlier signals without affecting the entire signal dataset (Yao et al. 2019), is adopted to design a signal pre-processor for cleaning sensor signals. The computing processes are below.

The median and the MAD (median absolute deviation) are important estimators of the Hampel filter. A set of signals is divided into multiple samples by a fixed size ($2K$) window, and the median of the sample is obtained below:

$$m_K = median\{x_{i,} \dots x_{i+K,} \dots x_{i+2K}\} \tag{1}$$

where m_K is the median of a moving data window; x_i is a sensor signal in the signal set; K is a positive integer called the half-width of the window.

The absolute deviation of the signals on both sides of the median is calculated, and the scale estimator of MAD is given below:

$$S_K = \alpha \cdot median(|x\ - m_K|) \tag{2}$$

where S_K is the scale estimator of MAD; α is the unbiased estimator of a Gaussian distribution, and $\alpha \approx 1.4826$; $j \in [i, i + 2K]$.

Spikes or outlier signals could be judged based on Equation 3. That is, if the difference between a signal data x_i and the median m_K is greater than $t \cdot S_K$, x_i is judged to be a spike or outlier signal and it should be replaced with the median.

$$H_K = \begin{cases} x_{i,} & |x_i - m_K| \le t \cdot S_K \\ m_{K,} & |x_i - m_K| > t \cdot S_K \end{cases} \tag{3}$$

where H_K is the Hampel filter; t is the scale factor.

Figure 4 shows the result of the Hampel filter applied to the vibration signals in the case study. In the original vibration signals under the 6th cutting cycle (shown in Figure 4(a)), some spikes can be observed. These spikes affect the accuracy of a prediction model. After the use of the

Hampel filter, signal spikes can be eliminated in the original signals. The processed signals are depicted in Figure 4(b).

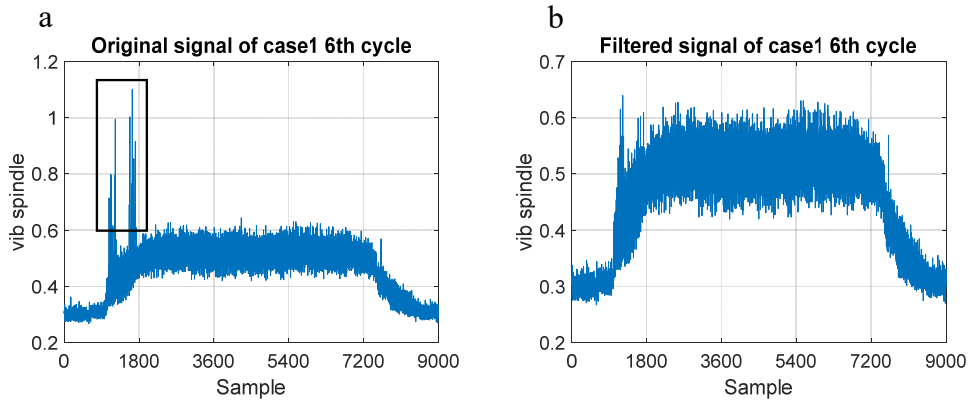


Figure 4.Signal processing using the Hampel filter: (a) Original data with spikes; (b) Processed data

3.2.3 Feature extraction

It will be time-consuming and error-prone to predict tool wear directly based on sensor signals. Thus, feature extraction is considered as a practical method to decrease the complexity of the prediction process. In this work, to ensure training accuracy, features under time-domain and frequency-domain are defined and extracted. Some considerations for the process are below:

According to Qin et al. (2019), statistic dimension features under time-domain might lead to a less accurate prediction accuracy as it could be affected by various manufacturing conditions, working loads and machine parameters. Instead, dimensionless features are more stable and sensitive to working conditions (Zhang et al. 2013; Sun et al. 2013; Hu et al. 2018), to compensate for the limits of dimension features. Hence, in this work, dimensionless and dimension features under time-domain are extracted to minimise the negative influence of working conditions and factors on prediction accuracy.

In the acquisition process of sensor signals, apart from the signals generated by a cutting tool, it is inevitably to collect signals generated from other elements/factors, such as signals from some abnormal machining conditions (e.g., looseness at the cutting tool joint), signals from the workpiece holder, etc. These signals are generally in different frequency bands (Nastac 2018). According to Wang et al. (2014) and Krishnakumar et al. (2015), changes in tool wear conditions can be reflected using features extracted from a specific frequency band in the spectrum, rather than from the entire band. In order not to miss critical features related to tool wear conditions, the frequency-domain of each sensor is equally divided into the low, middle and high-frequency bands, and features are extracted under those domains for optimisation and fine-tuning. The block diagram of the signal feature extraction is shown in Figure 5.

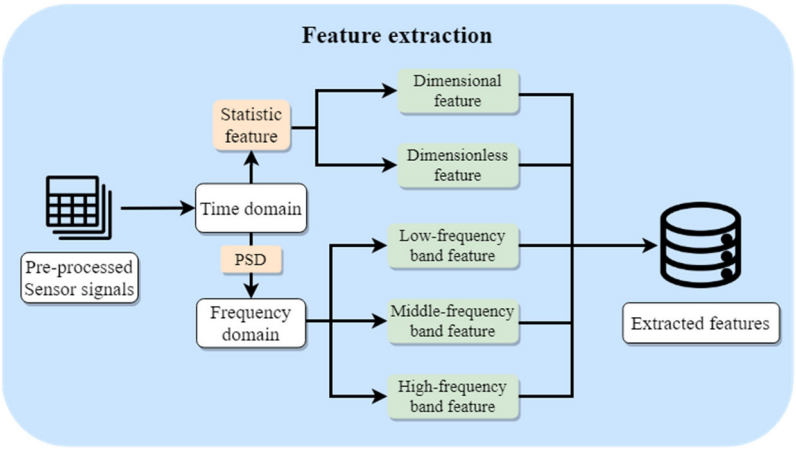
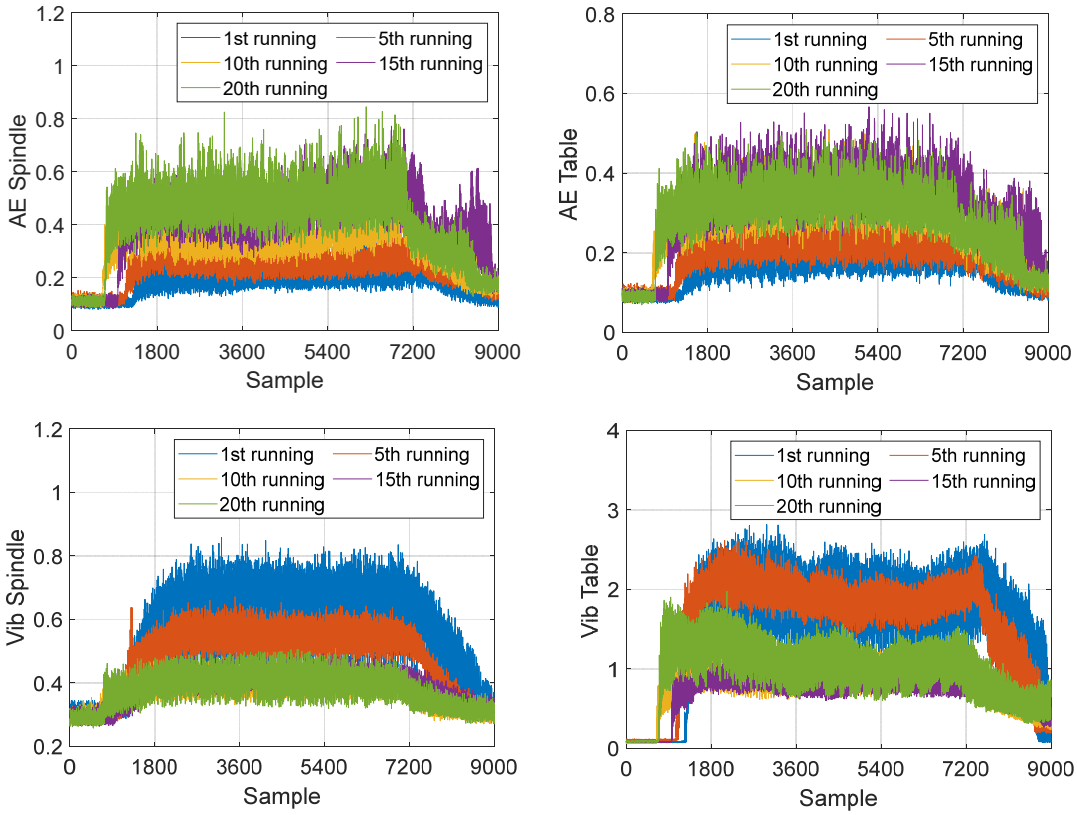


Figure 5.The block diagram for signal feature extraction

Feature extraction under the time-domain

The time-domain refers to the change of signal amplitude along time (Herff and Krusienski 2018). Signals used in this work were collected under time-domain. Take Case 10 shown in Table 14 as an example, signals of the AE, vibration and current under time-domain are shown in Figure 6.



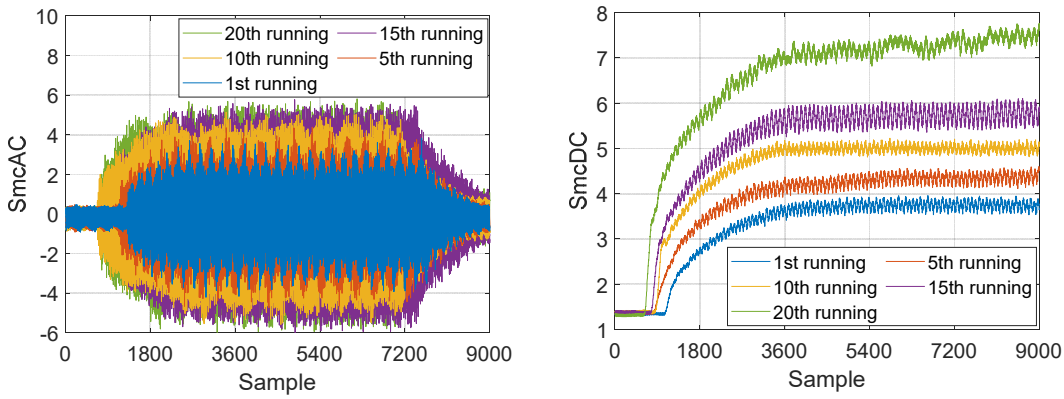


Figure 6. Signals of the Case 10 under the time-domain

It can be observed that, with the number of cuts increasing, the amplitude of the signals of current and AE rise. However, the vibration signal does not follow the same trend. As explained by Ahmada et al. (2015), machining causes the wear of cutting tools, the tool gradually becomes dull from the initial sharpness, and the contact region between the tool and workpiece is expanded during the process. Thus, the chatter of the cutting tool gets smaller, and the vibration amplitude drops. In addition, Bhuiyan and Choudhury (2013) drew a similar conclusion from the perspective of materials and cutting mechanism. As machining proceeds, the processing temperature and plastic deformation of the workpiece increase, resulting in reduced cutting force, which will eventually reduce the amplitude of the vibration signals as well. Moreover, plastic deformation is difficult to be detected by the vibration sensor in some cases, and the AE sensor is a better option. It is also considered as a reason that signals from multiple sensors are essential to be employed to reflect the overall machining trends and tool wear in a more comprehensive means. Signals for the rest cases display a similar trend presented above.

In this work, under time-domain, 11 dimension features and 6 dimensionless features are extracted from the signals of each sensor (Table 16). In total, there are 102 features under time-domain for all sensor signals (i.e., 17 features for each of the 6 sensors respectively).

Table 16. Features under the time-domain (Zhou and Xue 2018).

Time-domain			
Dimensional Feature		**Dimensionless Feature**	
Feature	**Formula**	**Feature**	**Formula**
Mean	$T_m = \frac{1}{n}\sum_{i=1}^{n} x_i$	Impulse factor	$T_{if} = \frac{\lvert T_{max}\rvert}{\mu}$
Max	$T_{max} = \max(x_i)$	Kurtosis factor	$T_{Kf} = \frac{T_{kur}}{{T_{rms}}^4}$
Min	$T_{min} = \min(x_i)$		
Standard deviation	$T_{std} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2}$	Margin factor	$T_{mf} = \frac{\lvert T_{max}\rvert}{T_r}$
Peak to peak	$T_{peak} = \lvert T_{max}\rvert - \lvert T_{min}\rvert$		
Root mean square(RMS)	$T_{rms} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} {x_i}^2}$	Shape factor	$T_{sf} = \frac{T_{rms}}{\mu}$
Skewness	$T_{ske} = \frac{\sum_{i=1}^{n}(x_i-\mu)^3}{(n-1){T_{std}}^3}$		
Kurtosis	$T_{kur} = \frac{\sum_{i=1}^{n}(x_i-\mu)^4}{(n-1){T_{std}}^4}$		
Mean absolute deviation	$T_{mad} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)$	Crest factor	$T_{cf} = \frac{\lvert T_{max}\rvert}{T_{rms}}$
Median	$T_{med} = x_{\frac{n+1}{2}}$ (odd dataset) $T_{med} = \frac{x_{\frac{n}{2}} - x_{\frac{n+2}{2}}}{2}$ (even dataset)	Skewness factor	$T_{skef} = \frac{T_{kur}}{{T_{rms}}^3}$
Variance	$T_{var} = \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}$		

In the table, $T_r = \left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\lvert x_i\rvert}\right)^2$ is the root value; $\mu = \frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert$ is the absolute mean value; x_i is a signal, *i*=1, 2, 3 ... *n*; *n* is the signal number.

Frequency-domain processing

Various uncertain factors may distort features under time-domain even after the denoising process on the signals, so that features only extracted under time-domain could be challenging to reflect tool wear properly. In this chapter, the power spectral density (PSD)-based method is designed for feature extraction under frequency-domain. PSD is a Fourier transformation of the autocorrelation function and describes the power of signals at different frequencies (Pappachan et al. 2017). The advantages of using PSD to analyse the frequency-domain are summarised as follows (Minbashi et al. 2016; Lee and Eun 2016; Xu et al. 2020):

- The irregularity of signals by the wavelength and amplitude can be displayed;
- Frequency distribution hidden in random signal noises can be revealed;

- Signal power caused by random changes can be minimised;
- PSD can help reduce spectrum leakage, prevent the signal from being disregarded, and avoid signals are mistakenly considered as a repeated period.

PSD can be obtained below:

$$F(k) = \sum_{T=0}^{N-1} f(t) e^{-2\pi i T \omega} \tag{4}$$

$$P(k) = \lim_{T \to \infty} \frac{|F(k)|^2}{T} \tag{5}$$

where $f(t)$ is a time-series signal; i is the imaginary unit and $i = \sqrt{-1}$; T is the time and $T = \{0, \cdots, N-1\}$; ω is the angular frequency; k is the spatial frequency, $F(k)$ is the Fast Fourier transformation; $P(k)$ is the PSD.

Abnormal behaviours of an object occur close to the natural frequency of the object (Chen et al. 2019). For a cutting tool, it will be more effective to capture features near the natural frequency of the tool for tool wear estimation. Features that are related to tool wear could be found in the sensitive frequency band of PSD. To promote the analysis, the frequency bands of sensor signals are divided into a low-frequency band (0-40 Hz), a middle-frequency band (40-80 Hz) and a high-frequency band (100-125 Hz).

The effect of using PSD is illustrated in Figure 7, where the PSD of the AE signals on the spindle for Case 11 is displayed. The graph is composed of three axes, i.e., frequency, flank wear and PSD. It shows that dominant frequencies happen in the low and middle-frequency bands. Moreover, the figure presents that the amplitude of each frequency band increases with the process of tool wear exacerbation. In particular, the PSD increment around 80Hz of the middle-frequency band is more intensive as the tool deterioration than that in the low-frequency band. It may attribute that the frequency of 80 Hz embodies the natural frequency of the system consisting of the cutting tool and workpiece. The increasingly severe friction between the two is reflected by the PSD.

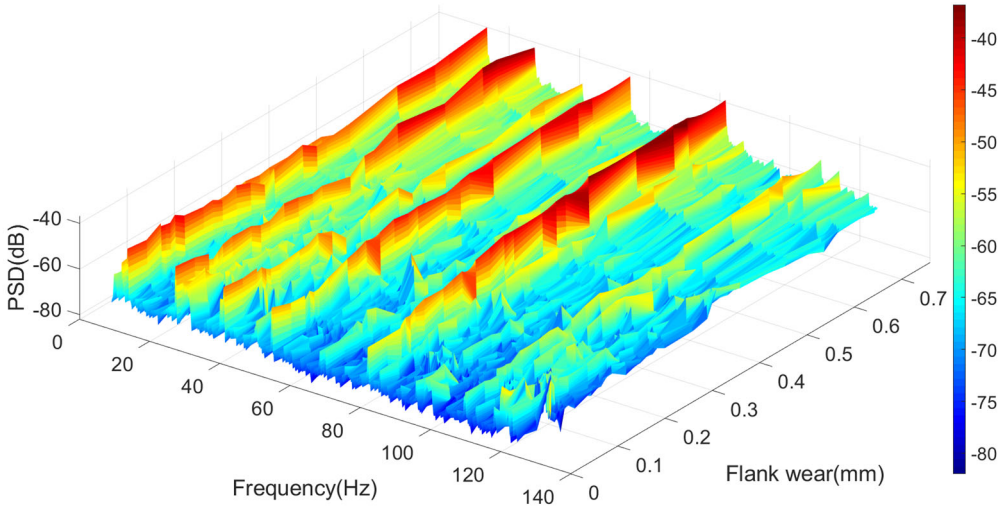


Figure 7. The PSD of the AE signals on the spindle

In this work, six features from the signals of each sensor are extracted for each frequency band. In total, 108 features are extracted for the six sensor signals under the three frequency bands. The related information is shown in Table 17.

Table 17. Features under the frequency-domain (Wang et al. 2015).

Frequency-domain	
Feature	**Formula**
Mean	$f_m = \frac{1}{n}\sum_{i=1}^{n} p_i$
Max	$f_{max} = \max(p_i)$
Min	$f_{min} = \min(p_i)$
Root mean square	$f_{rms} = \sqrt{\frac{\sum_{i=1}^{n} f_i^{\,2} p_i}{\sum_{i=1}^{n} p_i}}$
Frequency center	$f_{fc} = \frac{\sum_{i=1}^{n} f_i p_i}{\sum_{i=1}^{n} p_i}$
Root variance frequency	$f_{rv} = \sqrt{\frac{\sum_{i=1}^{n} (f_i - f_m)^2 p_i}{\sum_{i=1}^{n} p_i}}$

In the table, p_i is the power spectrum, i=1, 2, 3 … n; f_i is the frequency value; n is the spectrum line.

By performing the feature extraction, a total of 210 features are obtained from time and frequency-domains for the signals of the vibration, AE and current. Taking the features of the AE signals on the spindle for Case 10 as an example, some extracted features and flank wear measurement values are shown in Figure 8. It could be observed that the extracted features, such as mean and kurtosis under time-domain and RMS under middle-frequency-domain, are

consistent with the trend of tool wear, while others do not follow this trend. It reveals that features deviated from the trend of tool wear need to be identified and removed to improve prediction accuracy.

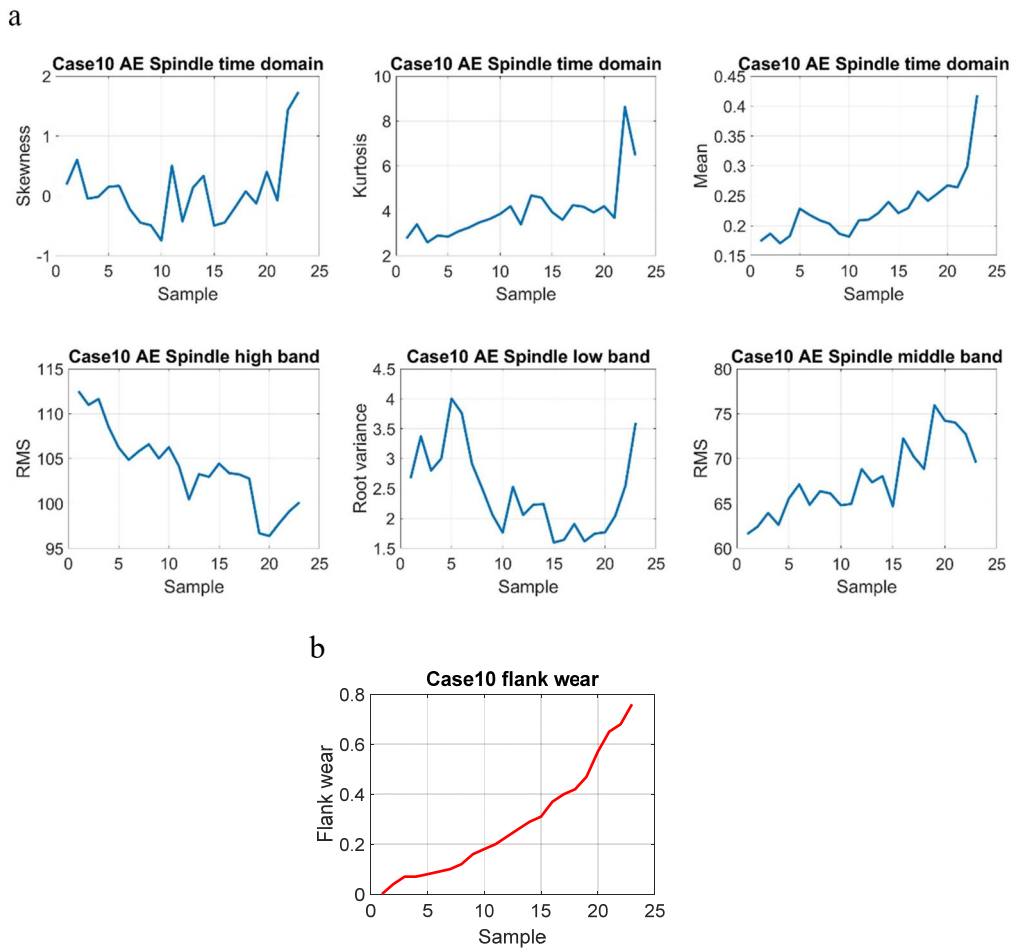


Figure 8. (a) Extracted features of the AE signals on the spindle, (b) Flank wear

Furthermore, to assess the effectiveness of the method proposed in this work, the PCC is adopted to analyse the linear relationship of the extracted features. The feasibility and necessity of this evaluation are that the quality of the input features has a direct impact on the prediction result. The correlation between features should be the smallest. Otherwise, a strong correlation will cause the CNN model to be biased when assigning weights. At the same time, these features with a high correlation will take additional calculations (Chicco and Rovelli 2019). The PCC is the value between [-1, 1] and can be calculated by the following equation:

$$P_{a,b} = \frac{Cov(a,b)}{\sigma_a \sigma_b} = \frac{\sum_{i=1}^{n}(a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_{i=1}^{n}(a_i - \bar{a})^2} \cdot \sqrt{\sum_{i=1}^{n}(b_i - \bar{b})^2}} \quad (6)$$

where, $P_{a,b}$ denotes the Pearson coefficient; $Cov(\cdot)$ denotes the covariance; σ denotes the standard deviation; a_i, b_i denotes the samples of each feature; $\bar{a}, \bar{b}$ denotes the mean value of each feature.

According to the coefficient value, the association between two features presents the positive or negative state, and a large value means a high correlation. Based on the features extracted above, their correlation matrix is shown in Figure 9.

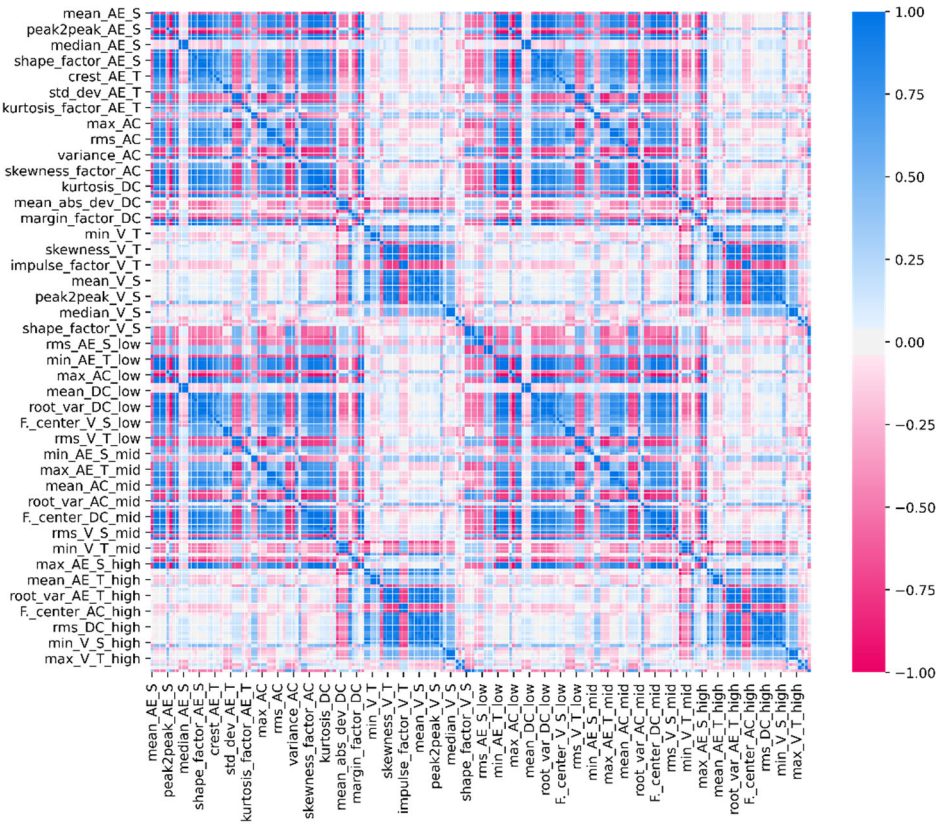


Figure 9. The correlation matrix of the extracted features

Figure 9 expresses the Pearson coefficients of the extracted 210 features by the shade of the colour. Based on the analysis, to obtain more efficient results for tool wear identification, feature selection, which to eliminate redundant or less relevant features, is a necessary process.

3.2.4 Selection of optimal feature

For extracted features, some features could be less important (relevant) to cutting tool conditions. It is worth identifying and removing those features to maintain the highest prediction accuracy. Moreover, the bigger the number of features, the more complex the prediction model may become, which will lead to lower computational efficiency.

In this work, the RFECV process is designed to select optimal features. On the basis of an SVM classifier, REFCV will work on a complete set of features to eliminate the least relevant feature recursively. That is, based on the importance score evaluations of features on tool wear conditions, REFCV is conducted to eliminate the least important feature in each iteration, to identify the optimal size of features for the highest accuracy of tool wear identification. Major steps of the SVMCV process is depicted below:

1. In view of the fact that a set of N features (a feature is denoted as x_i (i=1,…,N)) extracted from sensor signals could be in different ranges, a normalisation process is performed based on the Nadir and Utopia points to facilitate the following computations. The Utopia point z^U provides the lower bound of the features, and the Nadir point z^N provides the upper bound of the features. The normalisation process for each feature x_i (the normalised feature is denoted as x_i') is below:

$$x_i' = (x_i - z^U)/(z^N - z^U) \tag{7}$$

2. The set of normalised features are segmented into a training sub-set and a validation sub-set randomly according to an approximate 7/3 ratio for M times. Each randomly generated group is denoted as G_k (k=1,…,M). In each group G_k, a feature in the training sub-set in the group is represented as $T_{k,t}$ (t=1,…,p) and a feature in the validation sub-set is $V_{k,v}$ (v=1,…,q), where p and q are the numbers of features in the training and validation sub-sets respectively. Figure 10 shows a schematic diagram of this step.

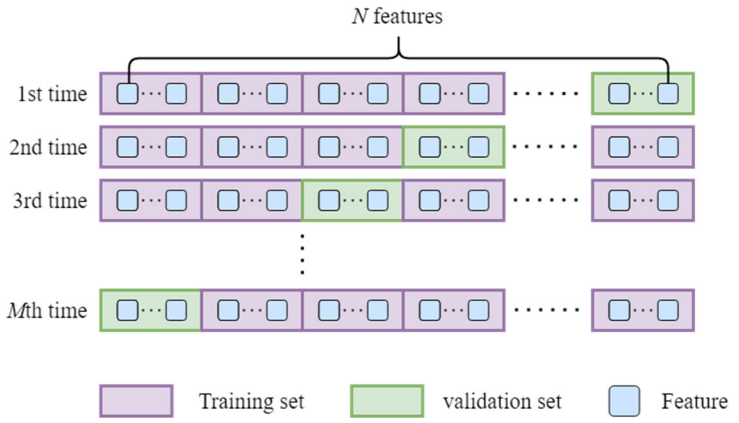


Figure 10. The schematic for the step 2

3. For each group G_j, the SVM classifier is used to conduct a binary classification process for tool wear (unworn or worn) based on its training and validation sub-sets according to the following procedures.

In the process, the SVM classifier distinguishes the unworn and worn statuses of features in the training sub-set through a hyperplane, and an optimal hyperplane will be achieved after the training. The maximum distance between features and the optimal hyperplane implies the lowest classification error (Nanda et al. 2018). In general, a hyperplane could be defined below:

$$y = \sum_{i=1}^{p} w_i * x_i' + b \tag{8}$$

where *y* represents the state of tool wear category ($y = 1$ means tool is worn, and $y = -1$ means tool is unworn), x_i' is a feature, w_i denotes a weight vector for the feature, and *b* is a constant of bias.

For the binary classification in this work, the plane for the two categories meets the condition of $H_0: w_i * x_i' + b \geq 1$ and $H_1: w_i * x_i' + b \leq -1$, which can be summarised as $y * (w_i * x_i' + b) \geq 1$. In addition, the solution of an optimal hyperplane that separates the categories with maximum distance is defined as $(w_i * x_i' + b)/\|w\|$. Thus, the problem can be transformed into the problem of $\max \frac{1}{\|w_i\|}$, which is equal to the following problem:

$$\begin{gathered} \min \frac{1}{2}\|w_i\|^2 \\ subject\ to: y * (\sum_{i=1}^{p} w_i * x_i' + b) \geq 1 \end{gathered} \tag{9}$$

In order to solve the above problem with inequality constraints, the SVM classifier further adopts a Lagrangian multiplier to transform the problem into an unconstrained form. At the same time, to eliminate the influence of nonlinear input features in the training process and improve the identification accuracy, the Gaussian kernel function $K(x_i', x_j')$ is employed to conduct feature mapping. In combining the above aspects, the optimisation problem for the SVM classifier is given below:

$$\begin{gathered} \min \frac{1}{2}(\sum_{i,j=1}^{p} \alpha_i \cdot \alpha_j \cdot y_i \cdot y_j \cdot K(x_i', x_j') - \sum_{i=1}^{p} \alpha_i) \\ subject\ to: \sum_{i=1}^{N} \alpha_i \cdot y_i = 0, 0 < \alpha_i < C, i = 1, 2, 3 \cdots N \end{gathered} \tag{10}$$

where α_i, α_j are the vectors for Lagrange multipliers, which correspond to a training sample $(x_i'\ y_i)$; x_i', x_j' are different features; y_i, y_j are different states of tool wear; $K(x_i', x_j')$ denotes the

kernel function, which indicates the mapping function for (x_i', y_i); C is the penalty coefficient, which is the tolerance for errors.

After the optimisation problem is solved, the weights of each feature are determined as Equation 11, and the average value of these weights is the weight of the SVM model trained by the training sub-set.

$$w = \sum_{i=1}^{N} \alpha_i \cdot y_i \cdot K(x_i) \tag{11}$$

Furthermore, based on the trained SVM model, the decision-making function of judging the status categories of the tool wear according to the new features can be represented below:

$$f(x) = sign(\sum_{i=1}^{N} \alpha_i \cdot y_i \cdot K(x, x_i) + b) \tag{12}$$

where $f(x)$ denotes category of the tool wear, $\in \{+1, -1\}$; x denotes a new feature.

After classifying the features using the validation sub-set through the SVM classifier, the identification accuracy is assessed by the confusion matrix. There are four possible results generated from an identification accuracy, i.e., true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). These four outcomes can be defined in the proposed system below:

- TP: The input signal feature indicates the cutting tool is unworn, and the cutting tool actual unworn.
- TN: The input signal feature indicates the cutting tool is worn, and the cutting tool actual worn.
- FP: The input signal feature indicates the cutting tool is worn, and the cutting tool actual unworn.
- FN: The input signal feature indicates the cutting tool is unworn, and the cutting tool actual worn.

The identification accuracy of the validation sub-sets can be assessed using Equation 13, which represents the proportion of the correct identification in the total identification.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

4. After each G_k (k=1,…,M) is processed according to Step 3, their average accuracy is used as the prediction accuracy of the feature subset with the size N (i.e., $\frac{1}{M}\sum_{i=1}^{M}(\text{Accuracy})$). Moreover, according to the calculated weight in Equation 11 and the importance ranking criterion of the SVM, $C_i = {w_i}^2$ (C_i denotes the importance score of Feature i), the importance of each feature in the size N can be ranked.
5. The feature with the least importance is discarded so that the number of features becomes N-1. The above Steps 1-4 are recursively repeated until all features are eliminated. During the process, the sets of features are denoted as S_N, S_{N-1},…, S_1.
6. For S_N, S_{N-1},…, S_1, their (average) prediction accuracies are ranked. The number of features in the set with the highest prediction accuracy is determined as the optimal number of the feature selection. According to the importance score and the optimal number of the features, optimal features are selected. The flowchart of the proposed methods is shown in Figure 11.

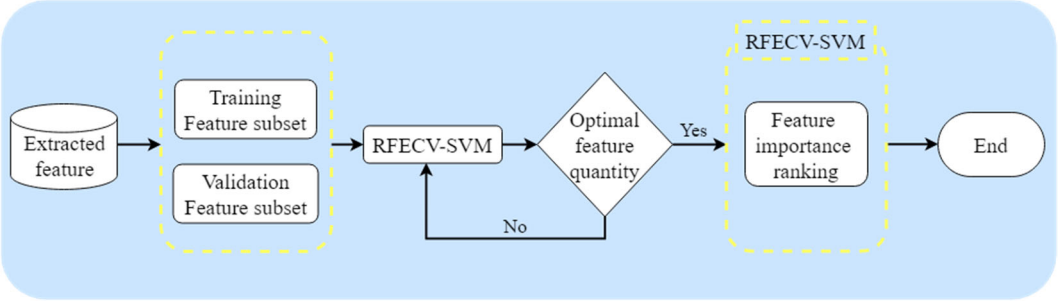


Figure 11. The flowchart of the RFECV feature selection

According to the above process of RFECV, the best feature subset of each sensor signal is selected. Taking the AE signal on the spindle as an example, the prediction accuracy of the feature subsets of different size under time and frequency-domains are shown in Figure 12 and Figure 13, respectively.

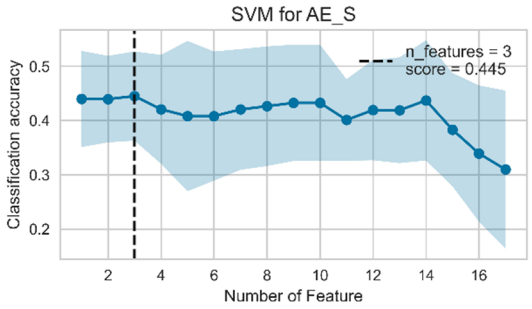


Figure 12. The classification accuracy based on time-domain features for the AE signals on the spindle

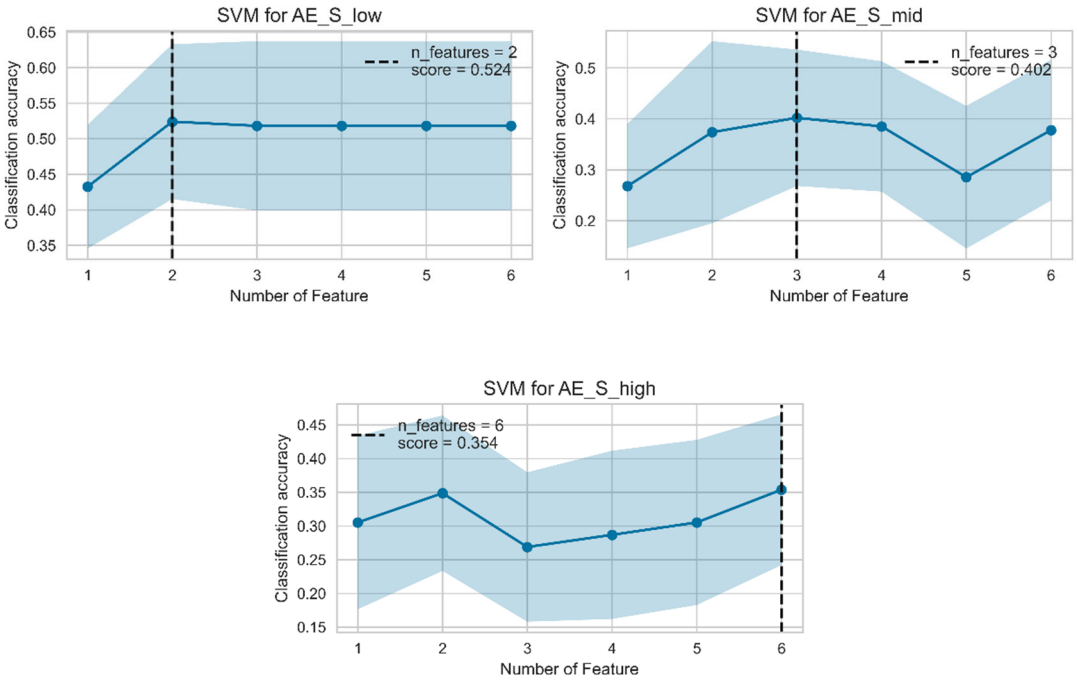


Figure 13. The classification accuracy based on frequency-domain features for the AE signals on the spindle

From Figure 12, the highest classification accuracy for the AE signal on the spindle is achieved by the number of features of 3. The optimal feature subset size of other signals under time-domain and frequency-domain are selected by the same method.

The contribution degree of each feature to the tool wear identification is sorted as shown in Figure 14 and Figure 15, and the features corresponding to the optimal size of the feature set can be selected.

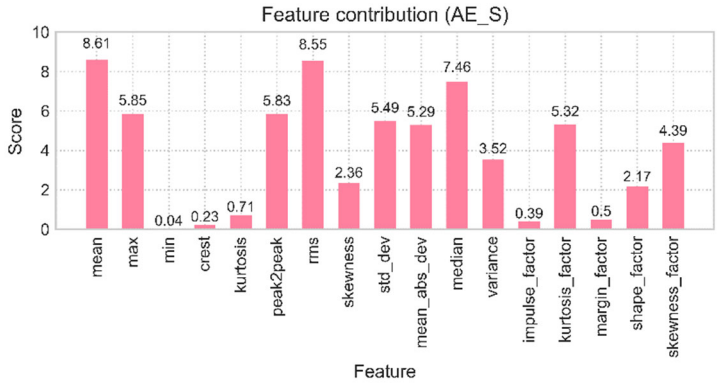


Figure 14. Importance ranking of features under the time-domain for AE signals on the spindle

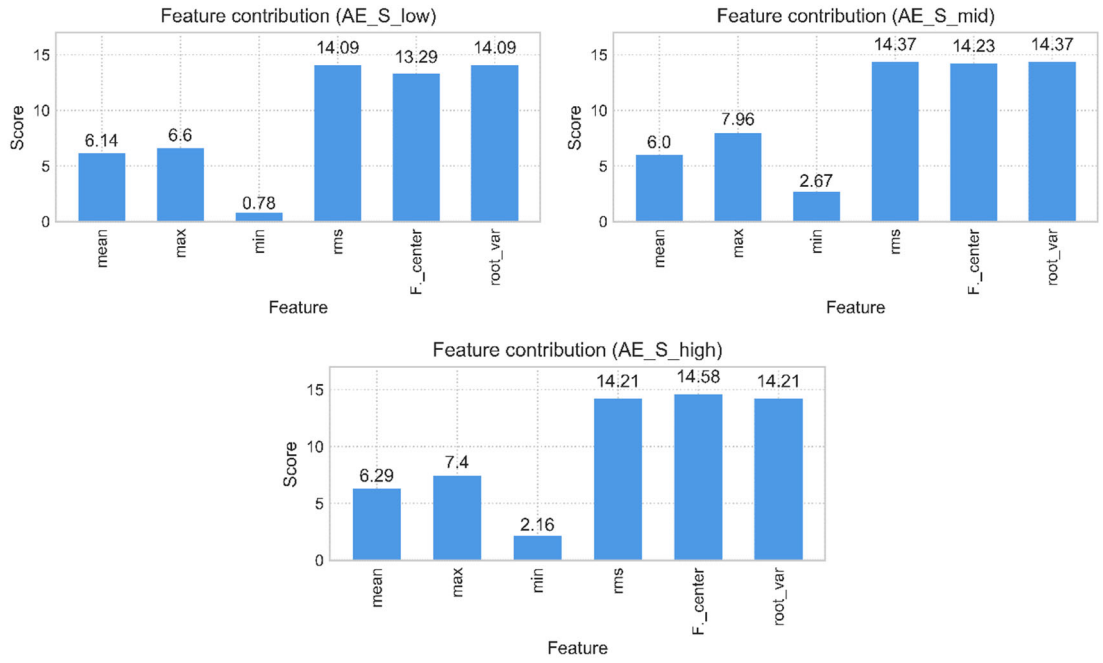


Figure 15. Importance ranking of features under the frequency-domain for the AE signals on the spindle

These scores indicate the importance of each feature to the tool wear status. It could be seen, under time-domain, the most relevant feature of AE signals on the spindle to the tool wear is mean, followed by RMS and median. The same information can be gained from different frequency bands in frequency-domain. At this point, the optimal feature selection for all signals is accomplished. There are 41 features under time-domain and 69 under frequency-domain are selected, as shown in Table 18 and Table 19.

Table 18. Optimal features under the time-domain

Time-domain		
Signal	**Feature number**	**Feature**
AE_T	5	RMS, Mean, Median, Standard deviation, Maximum
AE_S	3	Mean, RMS, Median
AC	10	Kurtosis factor, Impulse factor, Mean, Variance, Standard deviation, RMS, Crest factor, Minimum, Peak to peak, Mean absolute deviation
DC	10	Mean, RMS, Maximum, Median, Minimum, Kurtosis factor, Variance, Peak to peak, Skewness factor, Standard deviation
V_table	7	Shape factor, Kurtosis, Variance, Standard deviation, Skewness, Mean absolute deviation, Margin factor
V_spindle	6	Kurtosis factor, Variance, RMS, Mean, Median, Shape factor

Table 19. Optimal features under the frequency-domain

Frequency-domain						
	Low-band		Mid-band		High-band	
Signal	**Feature number**	**Feature**	**Feature number**	**Feature**	**Feature number**	**Feature**
AE_T	2	Max. Frequency center	2	RMS, Root variance	4	Max., Root variance RMS, Frequency center
AE_S	2	RMS, Root variance	3	RMS, Root variance, Frequency center	6	RMS, Max., Frequency center, Root variance, Mean, Min.
AC	4	Frequency center, Max., Mean, RMS	4	Mean, Min., Max., RMS	6	RMS, Mean, Max., Min., Root variance, Frequency center,
DC	4	Frequency center RMS, Mean, Max.	5	Mean, Max., RMS, Frequency center, Root variance	4	Frequency center, RMS, Variance, Max.
V_table	3	Frequency center, RMS, Root variance	3	Frequency center, RMS, Root variance	3	Frequency center, RMS, Root variance
V_spindle	5	Frequency center, Max., Mean, RMS, Root variance	4	Min., Max., Mean, Root variance	5	RMS, Root variance, Mean, Max., Frequency center

After the above process, the Pearson coefficient calculation based on the selected optimal features is carried out. Figure 16 displays the correlation matrix of the selected 110 features. The results show that, compared with Figure 9, the correlation of the features is reduced to some extent. That is, the average correlation is 0.23, which is much smaller than 0.47 of the extracted features. However, the correlations between some features still not eliminated in the above feature selection process. To further enhance the performance of the prediction model and simplify the subsequent training, these optimal features need to be further reduced. This process will be implemented by the dimensionality reduction based on Isomap in the following subsection.

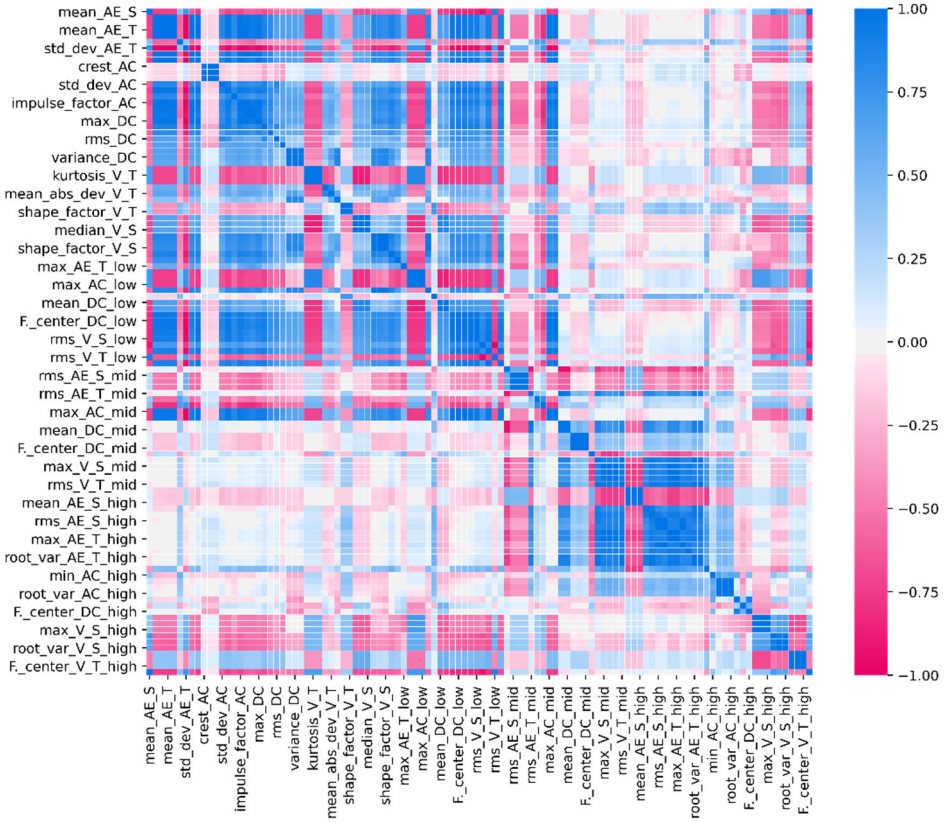


Figure 16. The correlation matrix of optimal features

3.2.5 Dimensionality reduction

In this work, selected features are further fused by a manifold learning dimensionality reduction method. Generally, the PCA is one of the most prevalent methods for data dimensionality reduction. PCA can transform features into uncorrelated values and reserve original information (Rad et al. 2014). However, a critical limit of PCA is that it is not suitable for data with nonlinear correlations, especially for high-dimension variables (Sahu and Nayak 2018; Caggiano et al. 2018; Kong et al. 2019). To address this issue, Isomap is employed in this chapter. According to Wang et al. (2017), Isomap extends the measurement of the multidimensional scale that hugely enhances the computing efficiency, global optimisation and information retention. Following the basic principle of the dimensionality reduction, the nonlinear dimensionality reduction is realised using Isomap through mapping high-dimensional data into a lower-dimensional space. It is also remarkable that, Isomap introduces geodesic distance in the calculation of the shortest distance instead of the Euclidean distance, which is the metric of the multidimensional scaling (MDS) algorithm.

To better explain Isomap, it is necessary to explain the core idea of MDS, which is to maintain the sample distance in a low-dimensional space the same as in the original high-dimensional space.

It refers to obtain the Euclidean distance matrix D between m samples in the original d-dimensional space. The element d_{ij} in the matrix represents the distance between samples x_i and x_j (e.g., features). To achieve dimensionality reduction, the Euclidean distance between the newly obtained samples in the d'-the dimensional space Z ought to equal to the distance in the original space, that is, $\|Z_i - Z_j\| = d_{ij}$.

However, in the real situation, data that requires dimensionality reduction is usually nonlinear. It is not feasible to use the Euclidean distance to map the distance between sample points. Thus, the Isomap method introduces the geodesic distance to represent the actual distance between sample points, and the shortest distance matrix of the n-dimensional features is embedded into a low-dimensional space, as well as reserving the nonlinear characteristic of the original data.

In addition to effectively achieving nonlinear dimensionality reduction, the Isomap can also simultaneously implement the fusion of features from the different sensor signals, by mapping the features to a shared low-dimensional space, and then the original features can be fused in this space (Hu et al. 2019). This shared space is the space after dimensionality reduction, which retains the shortest distance between the features in the original space. In addition, each column in the shortest distance matrix finally obtained by the Isomap is a new representative component generated after dimensionality reduction.

Based on the above principle, Isomap-based dimensionality reduction could be executed as follows:

1. Isomap composes a manifold with the m-dimension matrix D, which is formed by the selected features. It then determines the neighbour point of each feature on the manifold, and connects these neighbour points to form a neighbour graph. Such a neighbour graph is shown in Figure 17.

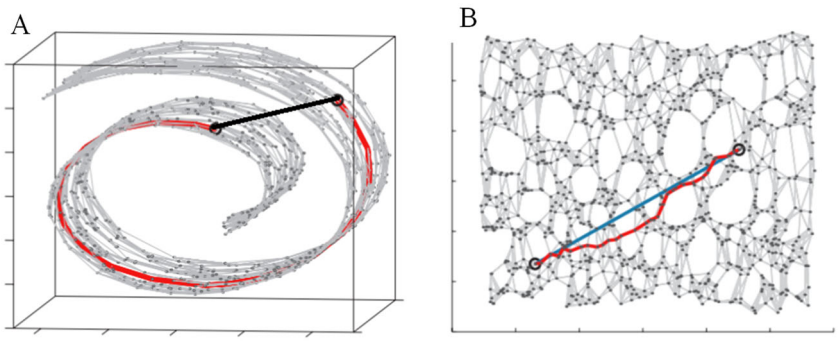


Figure 17. A neighbour graph

As shown in Figure 17 (A), the black line embodies the Euclidean distance between any two points. Obviously, it cannot represent the actual distance. And the red line indicates the actual distance between any two points along the manifold.

2. Calculating the geodesic distance between pairs of points in the neighbour graph to gain the shortest distance between any two points. This shortest distance matrix $D_{ij} = \{d(i,j)\}$ could be obtained by the Dijkstra's algorithm and Floyd–Warshall algorithm, where d is the shortest distance for any two points, i is the sample (feature), and j is its neighbour point. The shortest distance here denotes the newly generated feature vector based on the original feature set after the dimensionality reduction, and the feature vectors referred to as representative component in this work. As Figure 17 (B), the blue line is the shortest geodesic distance between the two points, and it is similar to the actual distance and can be used as an alternative.
3. Then, take D_{ij} as input to execute the MDS algorithm, to compute the inner product matrix B by applying the central matrix to D:

$$B = -\frac{1}{2}(d_{ij}^2 - \frac{1}{m}\sum_{i=1}^{m} d_{ij}^2 - \frac{1}{m}\sum_{j=1}^{m} d_{ij}^2 - \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} d_{ij}^2) \quad (14)$$

where, d_{ij} is the element in D_{ij}, represent the distance between sample i and j.

4. The lower-dimensional data matrix Z, which remains the most information of original feature set manifold and each column represents a new component feature, can be obtained by:

$$Z = \vee^T \wedge^{\frac{1}{2}} \quad (15)$$

where, $\vee$ is the eigenvector matrix of B, $\wedge$ is the diagonal matrix of eigenvalues of B.

5. According to the result of the cumulative variance, the different number of representative components are evaluated, and the component dimensionality with the largest variance is selected as the new feature set produced by the dimensionality reduction.

To avoid the overfitting problem caused by the CNN model training on a high-dimensional space and to improve computational efficiency, the Isomap method demonstrated above is adopted for the dimensionality reduction on the optimal feature set. According to the obtained shortest distance matrix, it is necessary to estimate the minimal number of components, which is the number of columns of the matrix, to describe the optimal feature set achieved earlier. With this aim, the cumulative variance sequentially accumulates the variances of each component to evaluate the proportion of the original information contained in the different dimensional features,

and this is the typical evaluation method for the dimensionality reduction (Mahecha et al. 2007). As shown in Figure 18, after the cumulative variance calculation, the first 10 components contain about 90% of the variance. That is, 90% of the raw information is preserved. When the number of components reaches 40, it can be used to describe nearly 100% of the information. Therefore, 40 components are finally employed as the latest features in the subsequent CNN model training.

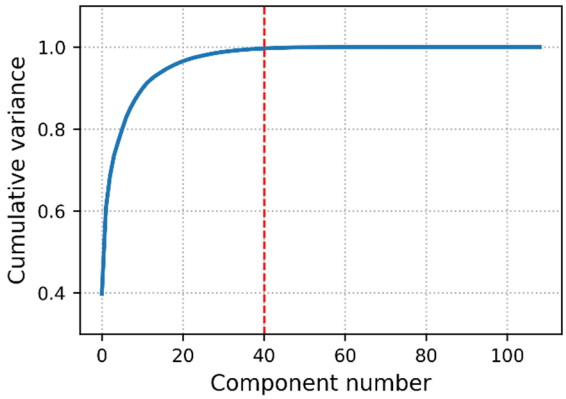


Figure 18. Cumulative variance of Isomap

In addition, with the obtained representative components, Figure 19 shows the correlation matrix of these 40 components. The correlation between these components has been greatly reduced, and their average correlation coefficient is only 0.025. Therefore, these components as a new feature subset are expected to bring higher performance to tool wear identification.

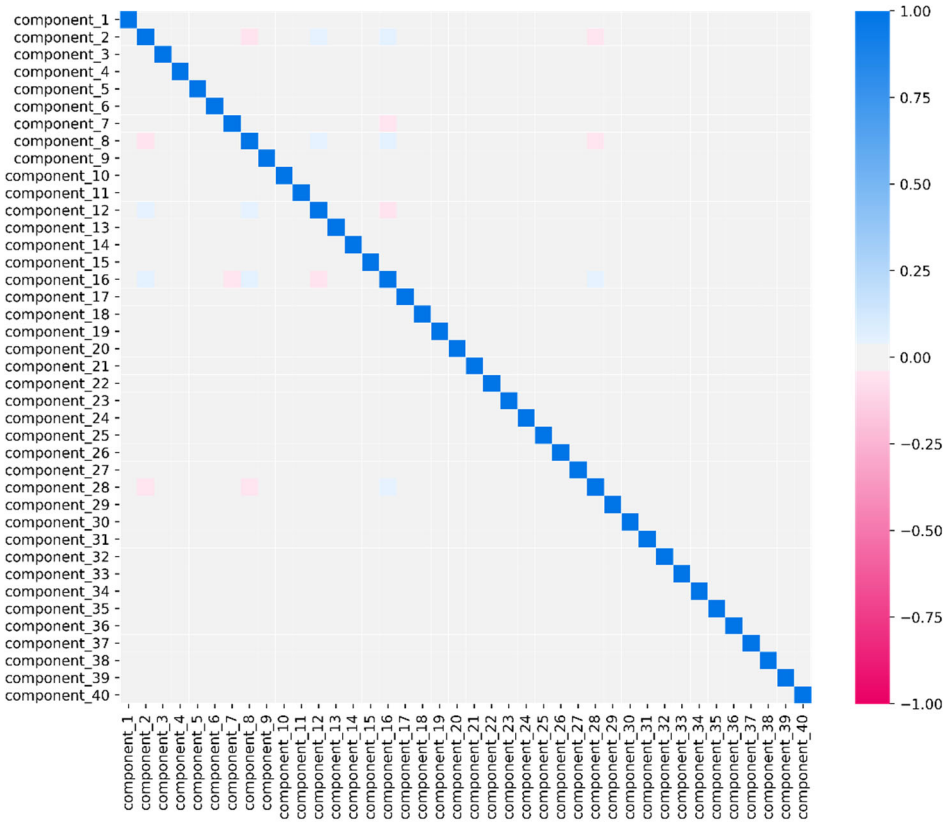


Figure 19. Correlation matrix of Isomap components

3.3 Prediction performance evaluation

3.3.1 Proposed system assessment

1D-CNN is adopted in the proposed system for tool wear identification. The basic architecture of the CNN model involves convolutional layers, pooling-layers and a fully connected layer. The convolution layer consists of trainable filters, and the filter will slide over the elements of the input data to generate a frame of the feature map in the next pooling layer. The depth of the convolution layer is the kernel number, which is defined based on the input data. A proper kernel size will increase the nonlinearity of the CNN model, and ensure the model to be suitable in processing complex data. While, a too-large kernel may affect computing efficiency and consume more memory (Soudani and Barhoumi 2019). A pooling layer after the convolution layer is adopted to reduce the dimension of input data and enhance the effectiveness of automatic feature extraction. The small size of a pooling layer window is often used considering a large size may cause the loss of a great deal of useful information (Yu et al. 2018). As the last component of the CNN model, a fully connected layer is employed for classification or regression on the features obtained from the previous convolution and pooling layers. Softmax is selected as an activation function in the fully connected layer. As summarised by Nwankpa et al. (2018), almost all of the fully connected layers of various deep learning models prefer softmax.

For the purpose of strengthening the performance of the CNN model for practical applications, an activation function is applied in the model. Commonly used functions are sigmoid, hyperbolic tangent (tanh), and ReLU (rectified linear unit). However, due to a large amount of computation, the sigmoid function may lead to the gradient disappearance, resulting in information loss or termination of the training of the CNN model. As an extension of sigmoid, the tanh function improves computational efficiency, but the problem of the gradient disappearance remains to be unsolved. ReLU is the most widely employed activation function (Nair and Hinton 2010). It is characterised by its faster computational speed, which makes the CNN model easy to optimise and to alleviate the occurrence of over-fitting problems (Ramachandran et al. 2017). It is generally in better performance than that of sigmoid and tanh for the CNN model. Therefore, ReLu is adopted for each convolution layer in the CNN model.

The following four feature sets were used for the validation of the CNN model for comparative analysis:

- The set of extracted 210 features,
- The set of optimal 110 features,

- The set of new generated 40 features (represented components) after the dimension reduction by Isomap,
- The set that is merged by the newly generated 40 features and 3 machining parameters, which are depth of cut, feed rate and work-piece material (as numerous studies show that machining parameters have significant influences on the tool condition (Saini et al. 2012; Xu et al. 2017; Dadgari et al. 2018), it also plays a positive role in tool wear identification).

The above four sets are fed into the 1D CNN model for the tool wear identification. The parameters and configuration of the CNN model are determined through trials and error comparisons. The fourth dataset input is used as an example. The kernel size of each convolution layer is 3. There is a max-pooling layer of size 2 after two convolution layers, which reduce the number of features and further the dimensionality for the next map. Before the next convolution layer, the dropout layer is applied to reduce interference between each sampling feature and prevent overfitting (Khodabandehlou et al. 2018). In the output layer, convoluted features are fully connected to be flattened to 80 neurons, then fully connected with the 32 neurons and 2 outputs, respectively (Figure 20).

Keras and Tensorflow are utilised to establish the CNN model. The CNN model is executed on the Apache spark, which is a cluster-computing framework (Bell 2014).

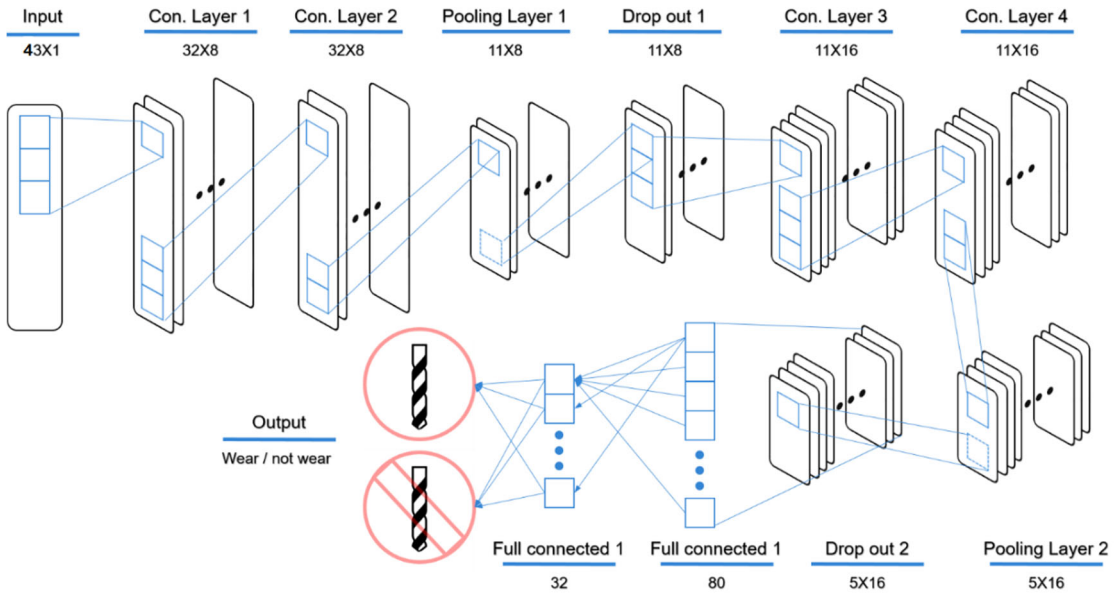
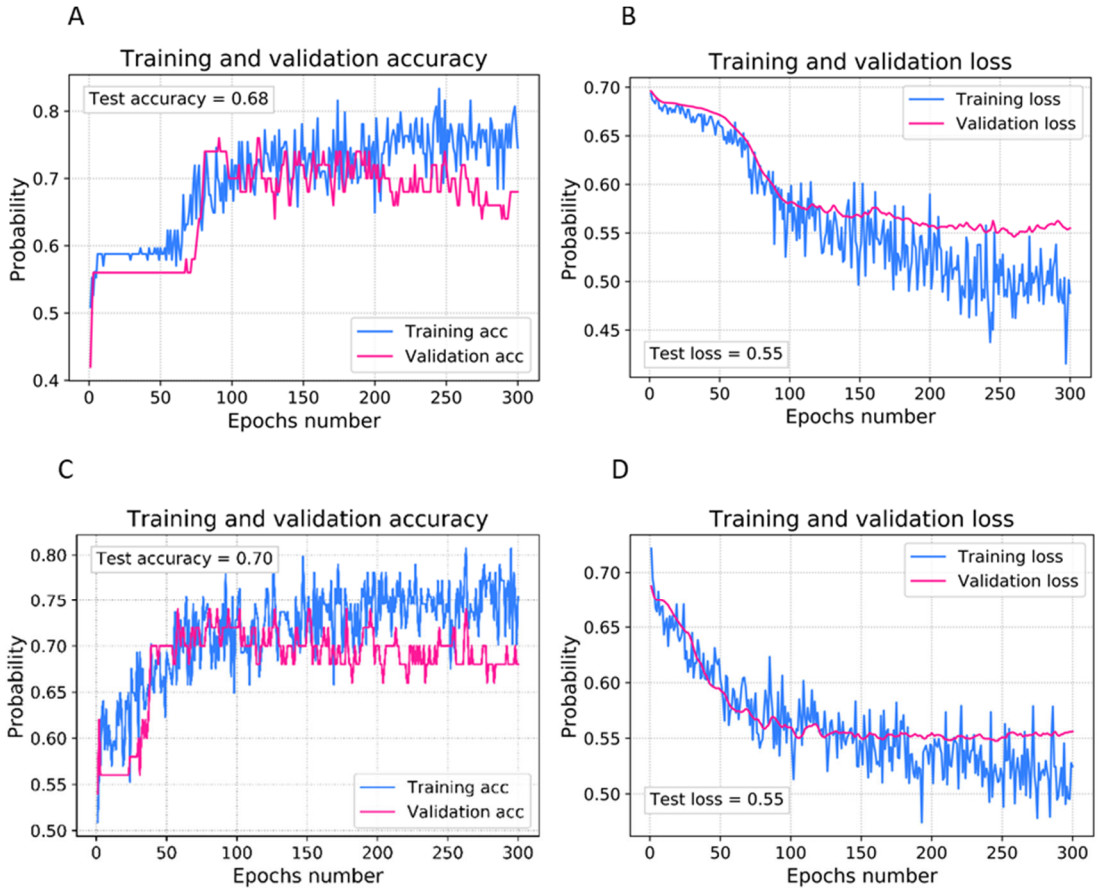


Figure 20. CNN model architecture

In the process of training and prediction, the architecture of the CNN model keeps unchanged to verify the performance of different subsets for the tool wear identification. The validation accuracy and validation loss are shown in Figure 21, it displays that the validation accuracy

increases along with the training accuracy until it reaches the optimum. Moreover, the validation accuracy of the four feature sets is summarised in Table 20. It provides the same outcome as Figure 21 that the accuracy of validation increases continuously with the adoption of feature selection and dimension reduction in sequence. The 4th feature set achieves the highest validation accuracy of 86%. Besides, it can be observed that the validation loss is reduced along with the training loss. However, it is not sustained in the first two feature sets, indicating that overfitting occurs. In general, the main reason for overfitting includes that the amount of data is too small, a large amount of noise information is included, and the prediction model is complex (Mehta et al. 2019). Since features are derived from the same raw dataset in this work, and the CNN model is unchanged, the small training loss of the last two sets proves that the proposed feature processing method effectively avoids overfitting by eliminating redundant information.



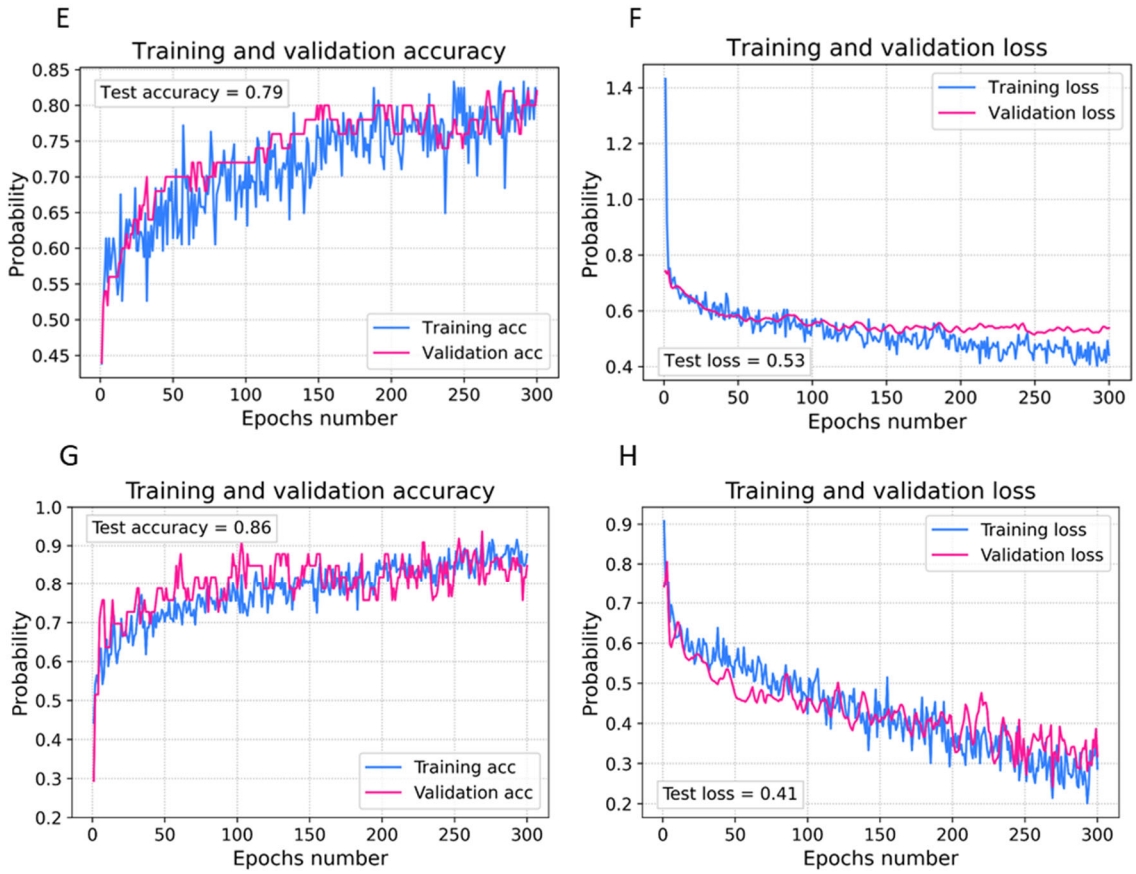


Figure 21. Accuracy and loss of the CNN model

Table 20. Data size and test accuracy of different input datasets

Input Data	Data size	Test accuracy
Extracted feature subset (Figure 21. A, B)	164×210	68%
Optimal feature subset (Figure 21. C, D)	164×110	70%
Isomap Component (Figure 21. E, F)	164×40	79%
Isomap Component and machining parameter (Figure 21. G, H)	164×43	86%

3.3.2 Comparison of SVM-based feature selection with other models

To verify the advantages of the presented feature selection method, comparative analyses between this method and other three prevalent machine learning algorithms, i.e., RF (Degenhardt et al. 2019), extra trees (ET) (Sharaff and Gupta 2019) and logistic regression (LR) (Kahya et al. 2020), are conducted. RF consists of a large number of decision trees. It uses the bagging method to train samples and assign weights to each of them to build a classification or regression model. It is also the ensemble of a large number of decision trees. ET adopts different methods to train the model, which is to train all samples at once and randomly divide the nodes of a single decision.

In addition, LR utilises the sigmoid function to estimate the classification probability between the feature and the dependent variable, and builds a binary classification model depending on this.

Based on the extracted feature in this work, three machine learning models, RF, ET, and LR are employed to perform the feature selection recursively. Thereby, the optimal number of features, the importance of each feature and corresponding optimal features are determined successively from 210 features. Figure 22 and Figure 23 show the partial results of the accuracy of the feature set of different size and feature importance ranking on the three models for the AE signals on the spindle (left side: time-domain, right side: frequency-domain).

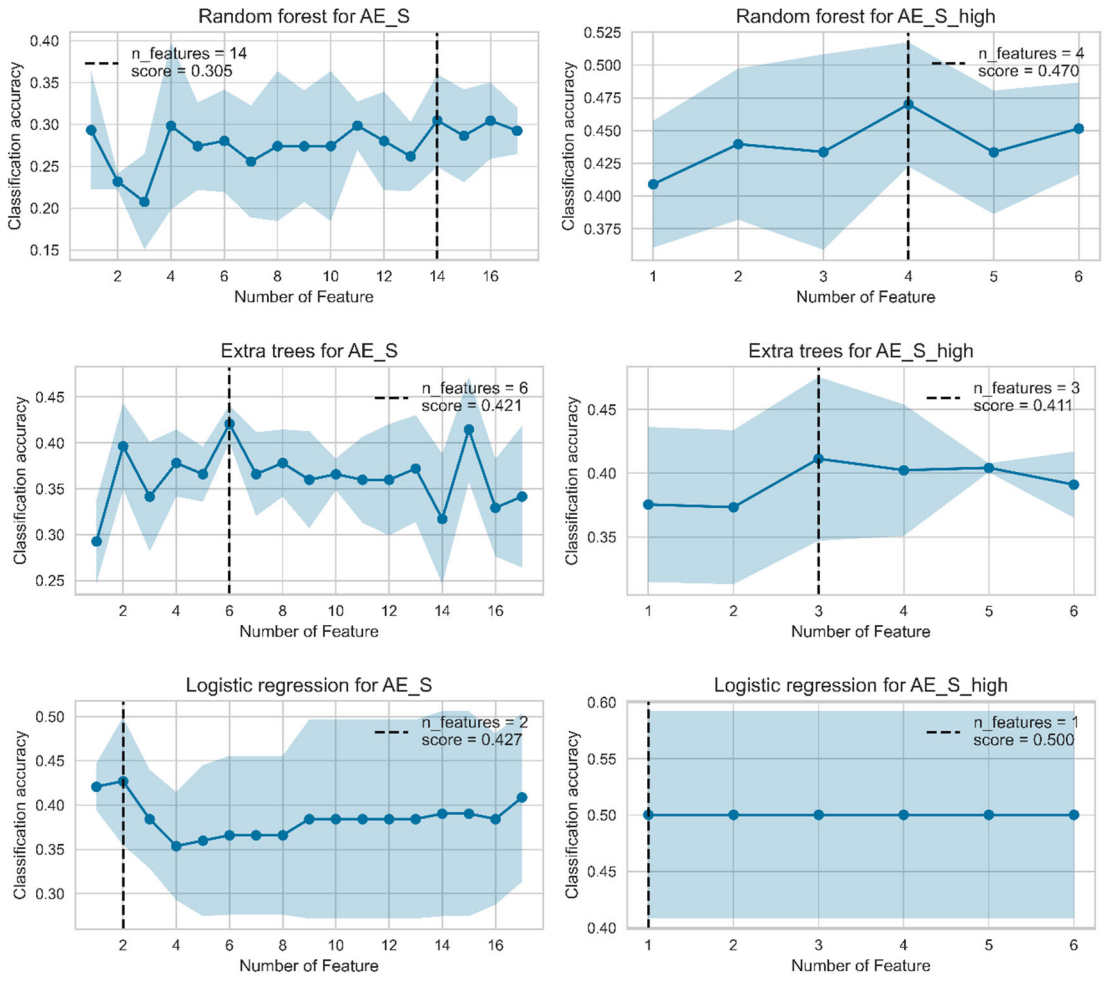


Figure 22. The classification accuracy of different feature sets determined by different models

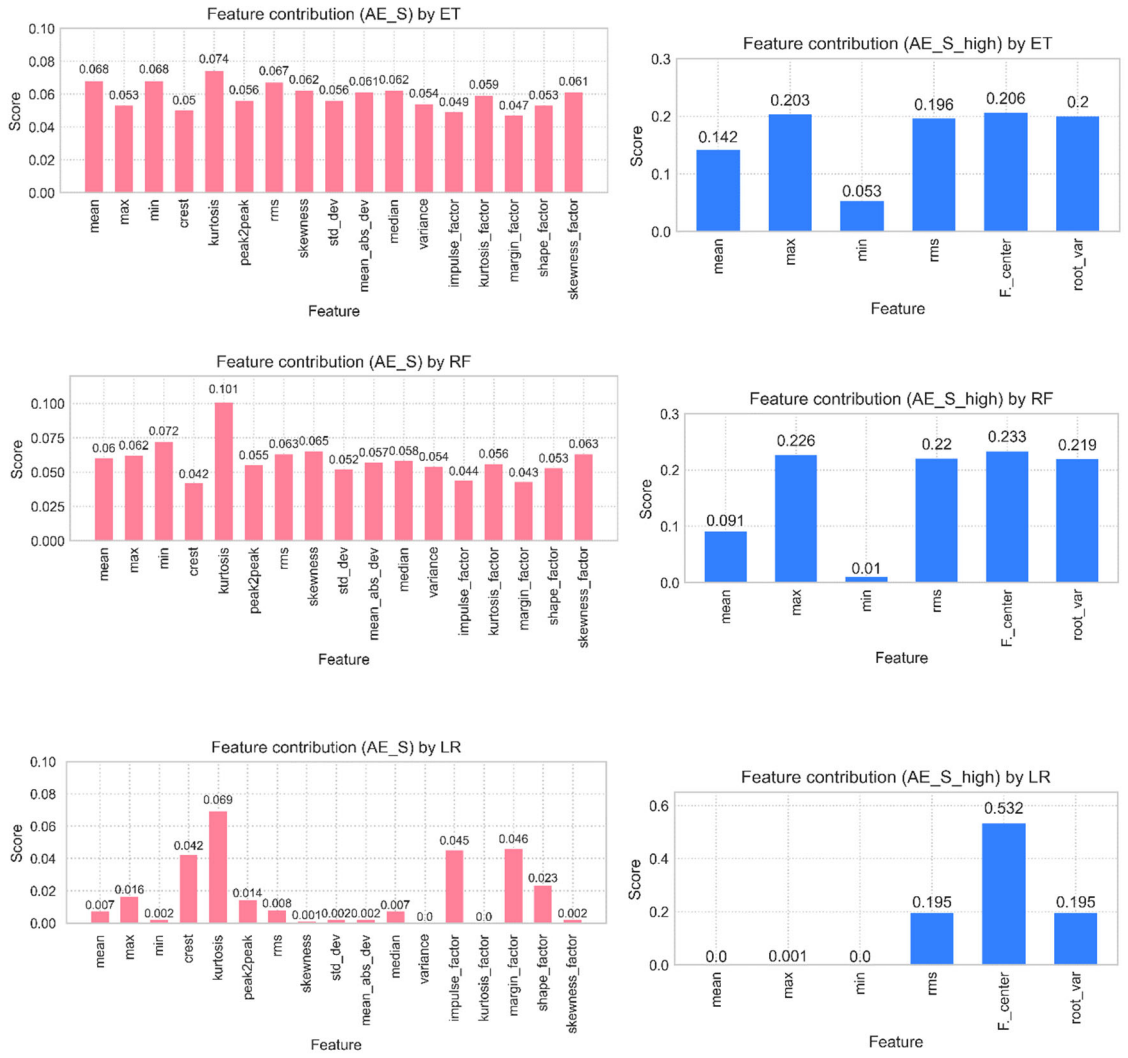


Figure 23. Importance ranking of features by different learning models

Different machine learning models have determined the different number of optimal features after performed the feature selection. The features selected under the time and frequency-domains are summarised in Table 21. It can be seen that, since the diverse performance of the three models, the evaluation results of the contribution of the size of the feature set and the features to the tool wear identification are not uniform. The number of extracted features is significantly reduced, and RF, ET and LR achieve the reduction of the data amount of 42%, 45% and 67%, respectively.

Table 21. The number of optimal feature selected by different learning models

Signals	Random forest				Extra trees				Logistic regression			
	Time-domain	Frequency-domain			Time-domain	Frequency-domain			Time-domain	Frequency-domain		
		low	mid	high		low	mid	high		low	mid	high
AE_S	14	2	4	4	6	4	3	3	2	3	3	1
AE_T	5	3	3	4	5	2	2	4	5	2	2	1
AC	10	4	4	6	10	4	4	6	6	4	1	1
DC	10	4	5	4	11	4	5	4	10	4	1	1
V_S	6	5	4	5	8	5	4	5	6	1	3	1
V_T	7	3	3	3	7	3	3	3	6	1	3	1
Total		122				115				69		

For the purpose of eliminating the high correlation between features and further fusing the features, Isomap-based dimensionality reduction is performed on the optimal feature sets of the three models. Figure 24 displays the cumulative variance calculation result of each model.

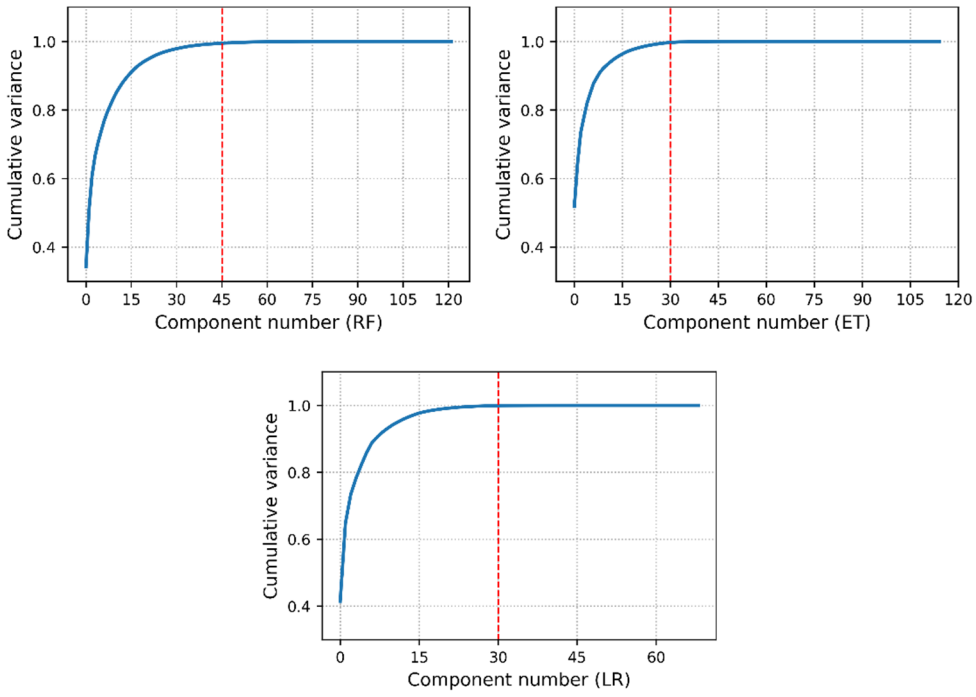


Figure 24. Cumulative variance of Isomap for different learning models

According to the cumulative variance, for RF, 45 representative components can effectively represent the source information of the selected optimal features close to 100%. To reach the same percentage, both ET and LR need 30 components. These newly generated features (components)

of RF, ET and LR model will be severally used for tool wear identification based on the proposed CNN model. Meanwhile, the feature selection with the RF, ET and LR achieved 78%, 85% and 85% of the data compression on the extracted features, respectively. In this regard, the proposed RFECV-based feature processing reduces the data volume by 81%. ET and LR in the four models have a slightly better dimensionality reduction effect.

Furthermore, Figure 25 shows the validation accuracy of the CNN model on three feature sets. Each feature sets contains the latest generated features (representative components) corresponding to the three comparative machine learning model, RF, ET and LR, and these feature sets were combined with the processing parameters as the input. It is found that, compared with the result in Figure 21, the prediction accuracy of features selected by RFECV-SVM is the highest, followed by RF is 77 %. The accuracy of ET and LR is only 74% and 72%.

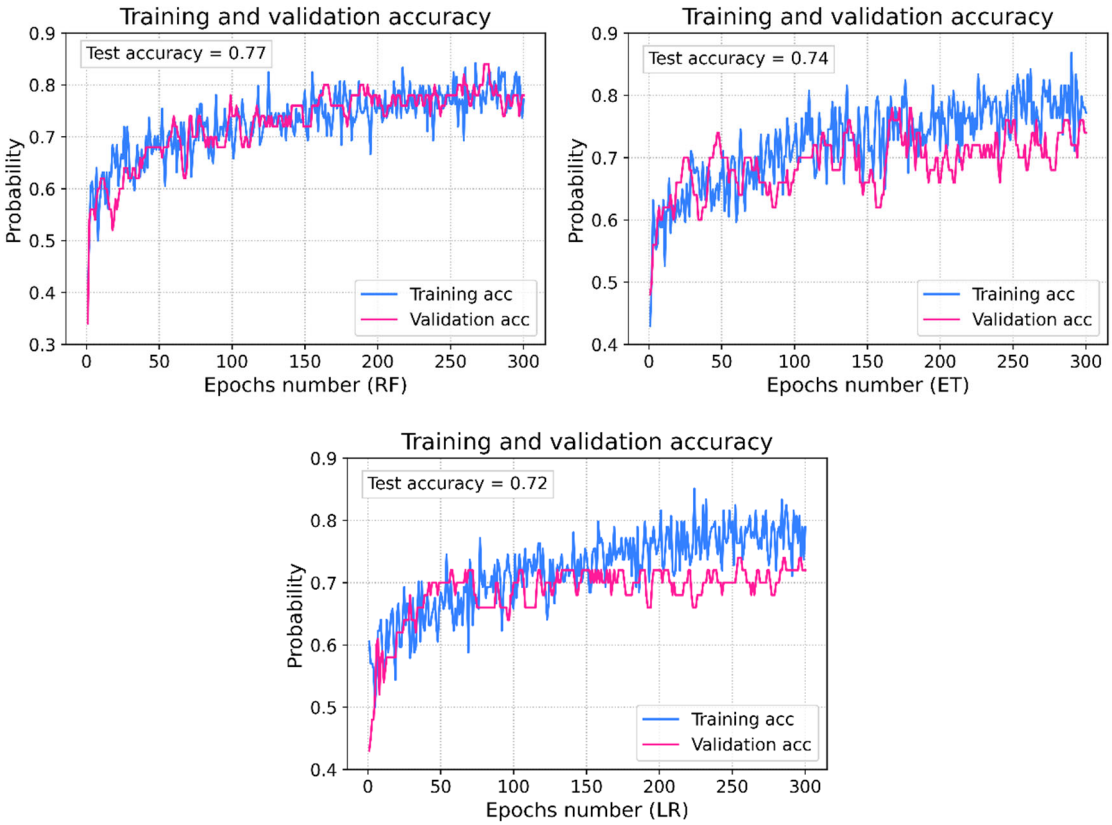


Figure 25. Validation accuracy with the feature subsets selected by different learning models

Apparently, considering both dimensionality reduction and prediction performance, the feature processing method of the RFECV-based feature selection and Isomap dimensionality reduction in the proposed system shows advantages. The obtained feature set could comprehensively describe the original multi-sensor signal, and provide robust performance for the tool wear identification.

3.4 Summary

In this chapter, a machining tool wear identification system has been developed based on multiple sensor signals. In order to reduce the data quantity and maximise the retention of data information, the proposed system applied the feature fusion technology, mainly includes optimal feature subset size selection, feature selection and dimensionality reduction. The multi-layer feature fusion model is developed based on the RFECV and Isomap method, to perform the deep processing and to eliminate the redundant signal of different sensors. With the support of feature selection and data fusion algorithm, this system emphasises the sensitivity of various sensor signal features to the tool wear, determines the optimal features of less correlation and ensures the accuracy of the subsequent prediction model. Moreover, to effectively alleviate the pressure of tool wear identification under multi-sensor signals and complex machining conditions, the system has employed the 1D CNN model to handle the classification task, and the simple structure generated the efficient prediction. In addition, the system executed the evaluation with a real dataset, which contains current, vibration and AE signals from actual machining. First of all, four feature subsets with the different number of features were obtained through the implementation of the proposed feature fusion method, they were then adopted as the input of the CNN model, and the result of the forecasting of tool wear is satisfactory, the optimal verification accuracy of 86%, which is can be improved by employing more sensor data, is achieved by the feature subset that after feature selection and dimensionality reduction. Meanwhile, the feature set size is reduced by 81%. Furthermore, the proposed feature processing have been compared with RF, ET and LR model, in terms of data volume reduction and tool wear identification accuracy, it proves that the proposed feature fusion method has competitive advantages, and the tool wear identification accuracy is guaranteed on the premise that the input data size is minimised.

Chapter 4. Tool RUL Prediction using Signal Partition and Hybrid Deep Learning

4.1 Introduction

The increasing complex processing environment of the cutting tools and the development of intelligent manufacturing strategies have made high dimensionality data, mainly, the multi-source sensor signals, displays apparent advantages to the tool RUL prediction. In order to quantify tool wear and achieve high-precision tool RUL prediction through multiple sensor signals, this chapter proposed a multi-channel hybrid CNN-LSTM deep learning model, which combines deep feature extraction, layer-based feature fusion and sequence regression efficiently. For the sake of enhancing the prediction accuracy to meet the industrial requirements, an innovative signal partition method based on Hurst exponent is developed in this chapter. The introduced method gives an effective solution to solve the issue of the sensor signals unbalanced distribution in different tool wear stages. Furthermore, the evaluation of the proposed hybrid deep learning model has been conducted in a real milling experiment. Moreover, the performance of the proposed system has been compared with some popular integrated deep learning models and signal processing method to assess the effectiveness.

The remaining part of the chapter is arranged as follows. The methodology of the proposed method is presented in Section 4.2. In the following Section 4.3 is the system validation result discussion. Finally, the summary is given in section 4.4.

4.2 Methodology and System Flow

The workflow of the system and the methodology for the RUL prediction on cutting tools are shown in Figure 26.

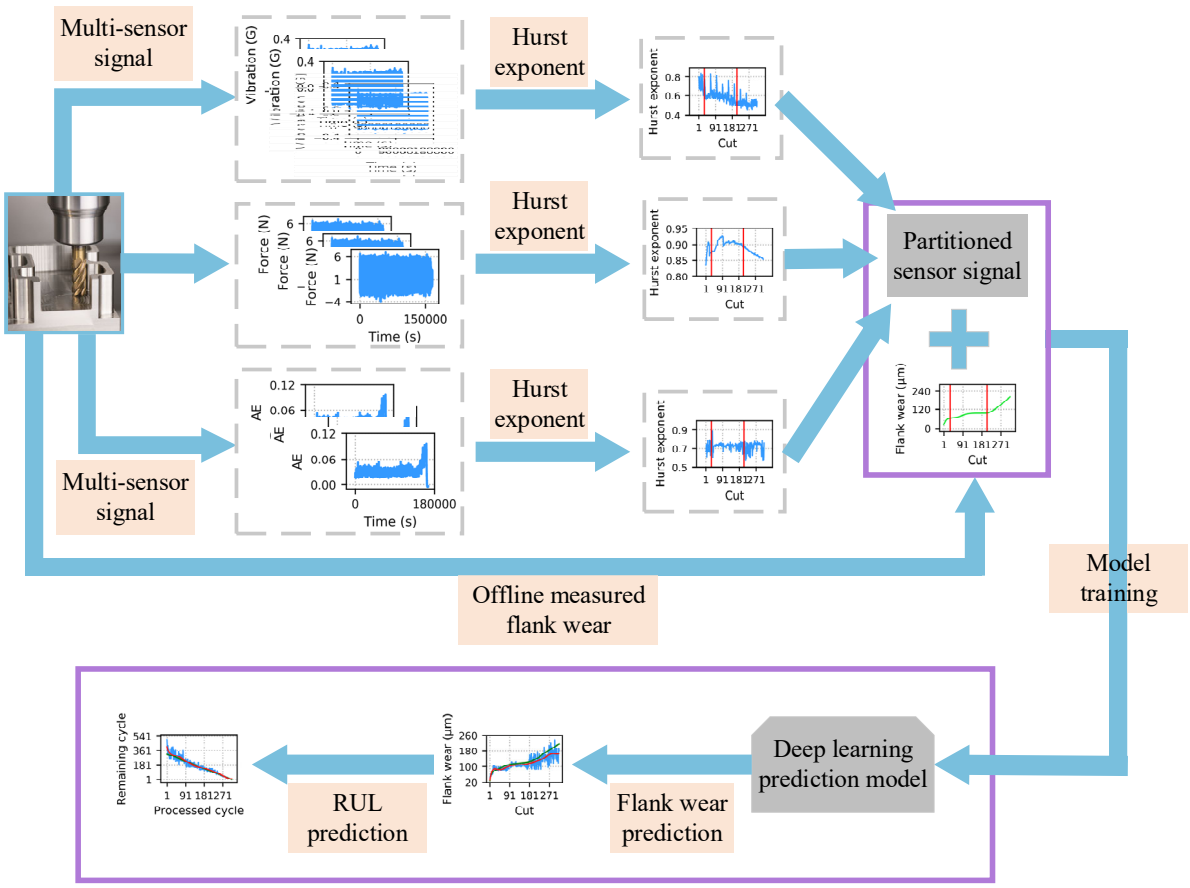


Figure 26. The flow and architecture of the presented system

The system consists of two subsystems, i.e., the Hurst exponent-based data partition, and a hybrid CNN-LSTM algorithm for RUL prediction. In the system, three types of sensors are deployed to monitor the statuses of a cutting tool, i.e., vibration (V), cutting force (F) and AE. The signal partition method based on the Hurst exponent is employed to segment sensor data into small batches to minimise the impact of feature imbalances from the input data. The partitioned datasets are then sent to the CNN-LSTM algorithm for processing, where the RUL prediction on a cutting tool is generated. In the CNN-LSTM algorithm, each CNN model is designed as a channel to process one type of signals, and the extracted features from the signals are fused in a concatenated layer, and the synchronised signals are further sent to the LSTM as a regression layer for predicting tool wear and RUL.

4.2.1 Signal partition based on the Hurst exponent

The input signals for the developed system are from multiple sensors. Multiple sensors V, F, and AE are mounted on a CNC machine. After a total *k* cuts are executed, each sensor collected *N* samples of data in each cut. The raw datasets $\{M^i | i = 1, 2, \cdots k\}$ are organised to obtain the sub-datasets for each sensor. Then, the Hurst exponent is adopted for partition on the data of individual cut from each sensor to acquire the input for the CNN-LSTM algorithm.

Signals from different sensors could have different impacts on the RUL prediction. The Hurst exponent is employed to process various sensor signals in order to establish accurate correlations between the signals and the prediction of tool wear. Key parameters of the Hurst exponent for sensor signals can be obtained using the following steps (Borys 2020):

For each window size *n* between N, N/2, N/4, N/8…until n approaching 350, repeat Step 1 – Step 6:

Step 1: Divide a given time-series of the sensor signal for each cut into M subseries of length n. Then calculate the mean value ($\bar{T}_m$) of the m^{th} subseries as follows:

$$\bar{T}_m = \frac{1}{n}\sum_{i=1}^{n} T_{i,m} \tag{16}$$

where, $T_{i,m}$ stands for the i^{th} signal in the m^{th} subseries and m = 1,2…M

Step 2: Create mean adjusted series D_m of the sensor signal $T_{i,m}$:

$$D_{i,m} = T_{i,m} - \bar{T}_m \quad for\ i = 1, 2 \ldots n \tag{17}$$

Step 3: Calculate the cumulative deviate series Z_m:

$$Z_{j,m} = \sum_{j=1}^{i} D_{i,m} \quad for\ i = 1, 2 \ldots n \tag{18}$$

Step 4: Calculate the range of R_m:

$$R_m = \max(\ Z_{1,m,}\ Z_{2,m}, \ldots, Z_{n,m}) - \min(\ Z_{1,m,}\ Z_{2,m}, \ldots, Z_{n,m}) \tag{19}$$

Step 5: Calculate the standard deviation, S_m, by using the following:

$$S_m = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(T_{i,m} - \bar{T}_m)^2} \tag{20}$$

Step 6: Calculate the rescaled range R_m/S_m and average over the all the M samples:

$$R/S = \frac{1}{M}\sum_{m=1}^{M} R_m/S_m \tag{21}$$

Step 7: Finally, the Hurst exponent of the sensor signal of each cut is obtained by linear fitting the log values of the rescaled ranges R/S for all the window sizes n based on:

$$\log(\frac{R}{S}) = \log C + H \cdot \log n \tag{22}$$

where C is a constant, H is the Hurst exponent.

The value of a Hurst exponent (i.e., H value) varies between 0 and 1, and it indicates the dependence of sensor signals on their past values (Lotfalinezhad and Maleki 2020). Based on the previous research, the following observations are made for various signals (e.g., V. F and AE in this chapter):

a) When $0 \leq H < 0.5$, there is a negative correlation between the sensor signals and flank wear. A smaller H value implies that the signal fluctuation in this period changes more dramatically.
b) When $H=0.5$, the corresponding signal presents the Brownian motion. It refers that the signal has no impact on future signals, and signal fluctuation will be completely random and unpredictable.
c) When $0.5 < H < 1$, the fluctuation of the signal shows a continuous positive correlation between the sensor signals and flank wear. The future trend will follow the changes of the present signal, and this type of serial signal is predictable.
d) When $H=1$, time-series signals will be in a straight line, and there is no fluctuation and correlation.

Within the lifecycle of a cutting tool, the value of a Hurst exponent is obtained from each signal by executing the above steps. The different values of the Hurst exponents will be used to establish the correlation between signals and various flank wear stages of a tool lifespan, which are normally divided into an initial region of rapid wear, a steady region of uniform wear and a severe region of dramatically increased wear. Accordingly, the sensor signals can be segmented based on the stages to facilitate the following prediction using a deep learning algorithm.

4.2.2 A hybrid CNN-LSTM algorithm for prediction

CNN has demonstrated exemplary performance at intelligent acquiring features from data and is immune to the frequency variation in the data. LSTM is more potent for time-series data learning. However, both algorithms present their restrictions when dealing with real-time data.

To leverage the superiority of both algorithms, in this chapter, a hybrid CNN-LSTM algorithm is designed for better prediction on tool life. It is capable of identifying the spatial and temporal relations in sensor signals. Based on the different types of signals, multiple CNN sub-models are constructed as pre-processors. The CNN sub-model parallelly extracts the features from each sensor node, and these features are then fused at a concatenate layer. Finally, the concatenated features from sensor signals are sent to the LSTM for prediction. The proposed CNN-LSTM model is illustrated in Figure 27.

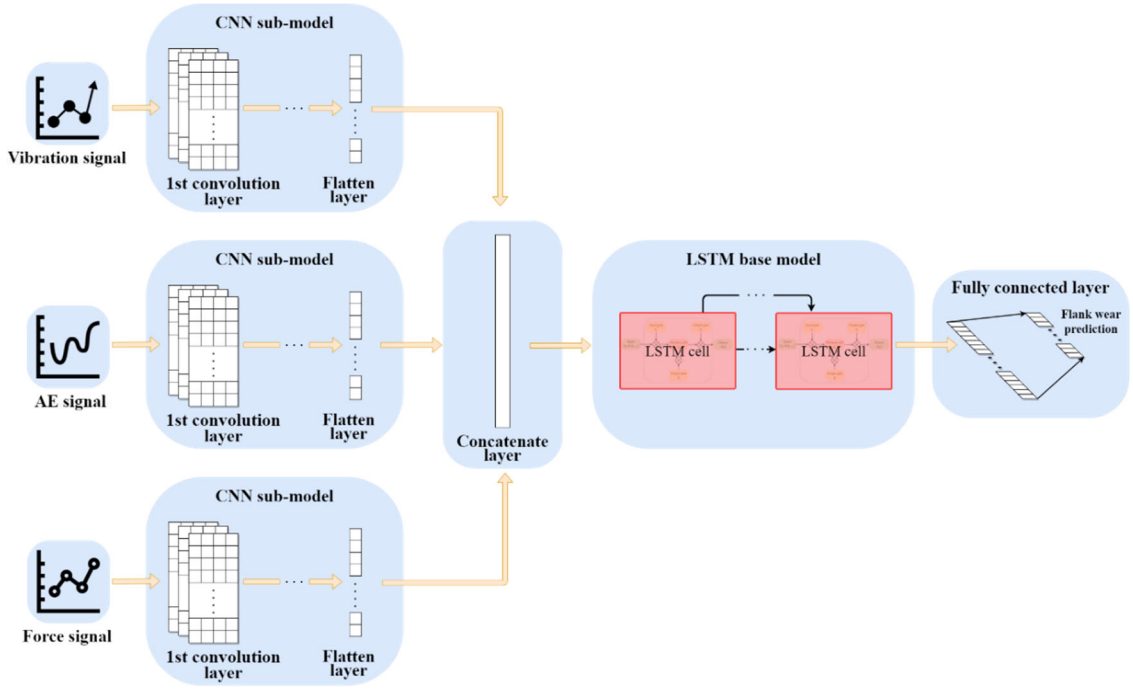


Figure 27. The architecture of the multiple channel hybrid CNN-LSTM model

CNN design

As introduced in the literature review chapter, CNN is one of the most potent deep learning models, offers the strength of extracting high-level dependency features automatically from input data. The learning ability and the training time of the CNN model are decided by its structure, especially the number of layers. Usually, a shallow structure cannot provide good processing performance, and meanwhile, an excessive deep CNN may be harmful to the time-sequential aspect of the data or cause overfitting (Chen et al. 2019, Zhu et al. 2020). 1D CNN better utilises its characteristic of automatic feature extraction and avoids the deviation. Moreover, based on the prediction model architecture established in this chapter, the subsequent regression prediction of LSTM seeks the time-series data as the input, rather than other formats, to preserve the temporal integrity of the sensor signal.

In the light of the advantages of 1D CNN, a 1D CNN structure is adopted in this chapter. In order to achieve a satisfactory prediction accuracy and efficiency in processing the sensor signals, the 1D CNN model with different numbers of convolution and pooling layers are evaluated to identify the most suitable architecture. The proposed model and several other models are summarised in Table 22. Based on the same input dataset (the exact dataset is described in Section 4.3), the prediction accuracies of each 1D CNN model are shown in Figure 28.

Table 22. The architecture of the adopted CNN models

Model	**Layer**					
	1st convolution	2nd convolution	3rd convolution	1st pooling	4th convolution	2nd pooling
A	●	-	-	●	-	-
B	●	-	-	●	●	●
C	●	●	-	●	-	-
D	●	●	●	●	-	-
E	●	●	●	●	●	●
Proposed	●	●	-	●	●	●

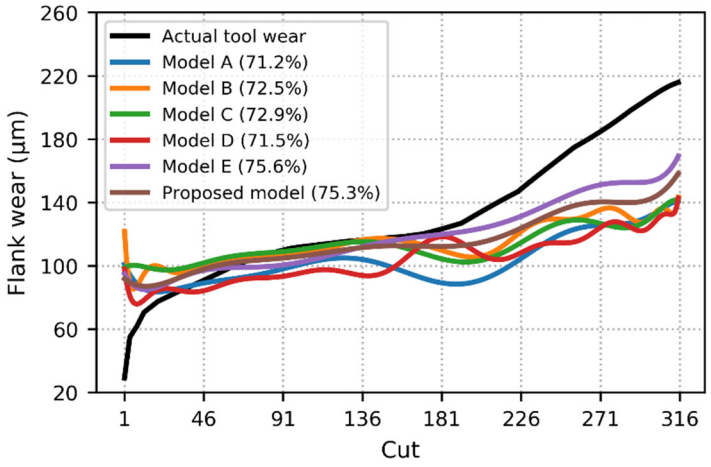


Figure 28. The prediction accuracies of the CNN models

Based on the above results, it clearly shows that, within the six 1D CNN architectures, the best prediction accuracy was achieved by the model E, which is 75.6%. The accuracy of the proposed model of three convolutional layers and two pooling layers is 75.3%, which is slightly lower than that of the model E. However, model E with the six-layer structure requires much longer computing time. Thus, the proposed model provides the best trade-off in accuracy and efficiency performance among the six models. The exact configurations of the proposed model are shown in Figure 29. To process three types of sensor signals, i.e., V, F and AE, there are three such CNN

models arranged in parallel (i.e., the CNN models of 1, 2 and 3 for the signals of V, F and AE respectively).

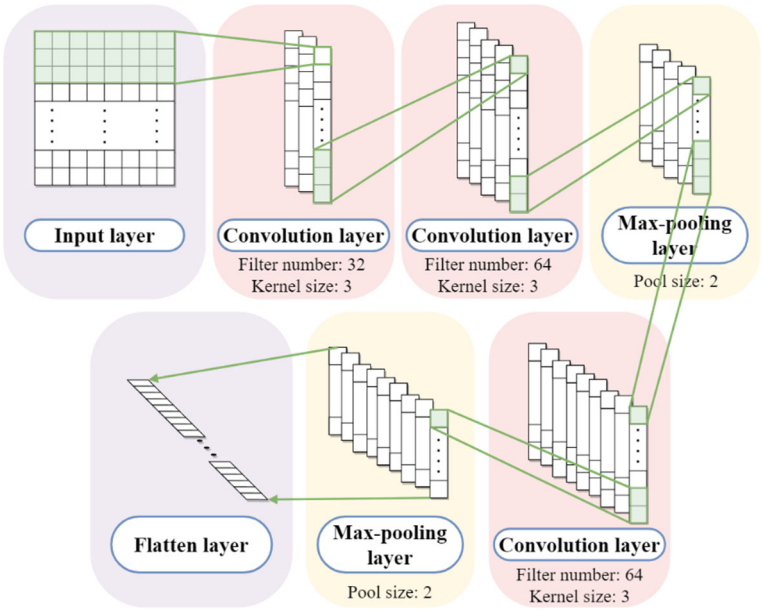


Figure 29. The architecture of the proposed CNN model

Each partitioned sensor signal in a format of the matrix is fed into the CNN model through its input layer. For segmented sub-matrix for the inputs of the CNN models of 1, 2 and 3, they are below:

$$A = \begin{pmatrix} a_1^1 & a_2^1 & a_3^1 & \cdots\cdots & a_n^1 \\ a_1^2 & a_2^2 & a_3^2 & \cdots\cdots & a_n^2 \\ & & \vdots & & \\ a_1^d & a_2^d & a_3^d & \cdots\cdots & a_n^d \end{pmatrix} \quad B = \begin{pmatrix} b_1^1 & b_2^1 & b_3^1 & \cdots\cdots & b_m^1 \\ b_1^2 & b_2^2 & b_3^2 & \cdots\cdots & b_m^2 \\ & & \vdots & & \\ b_1^d & b_2^d & b_3^d & \cdots\cdots & b_m^d \end{pmatrix} \quad C = \begin{pmatrix} c_1^1 & c_2^1 & c_3^1 & \cdots\cdots & c_e^1 \\ c_1^2 & c_2^2 & c_3^2 & \cdots\cdots & c_e^2 \\ & & \vdots & & \\ c_1^d & c_2^d & c_3^d & \cdots\cdots & c_e^d \end{pmatrix} \tag{23}$$

where, n, m, e denotes the number of signals of sensor A, B and C, respectively, d denotes the number of cuts in the dataset after segmentation.

The convolution layer of each CNN model convolutes the input matrix to generate the spatial feature map by the activation function, and it can be described as:

$$\Phi_l = f(\mathrm{conv}(X_{l-1} * \mathrm{w}_l) + \mathrm{b}_l)) \tag{24}$$

where, the Φ_l denotes the feature map of the lth convolution layer, X_{l-1} denotes the generated feature map from $(l-1)$ th layer, w_l denotes the weight, b_l denoted the bias, the $\mathrm{conv}(*)$ denotes the convolution process. The $f(\cdot)$ denotes the activation function. The ReLU is

selected as the activation function for each convolution layer in the proposed CNN model. The reason is that the ReLU increases the nonlinearity between layers, and only a small amount of computation is needed to alleviate the gradient vanishing problem (Carneiro et al. 2016).

After the convolution layer, the pooling layer is applied to further reduce the feature dimensionality, the max-pooling function is selected for every pooling layer, which extracts the maximum feature value with a window of size 2, so as to retain the important feature information and improve training efficiency. The output feature map can be depicted as:

$$\Phi_k = \mathrm{w}_k \cdot \max(\Phi_l) + \mathrm{b}_k \tag{25}$$

where, the Φ_k denotes the feature map of kth pooling layer, w_k denotes the weight, b_k is the bias, max $(\cdot)$ denotes the max-pooling function.

Finally, the flatten layer is connected to the last pooling layer, to complete the transition from the convolution layer to the next layer base model by converting the generated features into a one-dimensional array. And the size of the output array equals the number of the cuts d.

So far, the signal features extracted by the CNN model eliminates the interference between the heterogeneous sensors, and these features can be fused without complicated operations. Therefore, the concatenation layer is applied to be responsible for the feature fusion after the multi-channel CNN models. As the element number of each feature vector is d, the feature fusion can be depicted as:

$$\begin{bmatrix} \mathrm{a}'_1 \\ \mathrm{a}'_2 \\ \vdots \\ \mathrm{a}'_d \end{bmatrix} + \begin{bmatrix} b'_1 \\ b'_2 \\ \vdots \\ b'_d \end{bmatrix} + \begin{bmatrix} c'_1 \\ c'_2 \\ \vdots \\ c'_d \end{bmatrix} \Rightarrow \begin{bmatrix} \mathrm{a}'_1 & b'_1 & c'_1 \\ \mathrm{a}'_2 & b'_2 & c'_2 \\ \vdots & \vdots & \vdots \\ \mathrm{a}'_d & b'_d & c'_d \end{bmatrix} \tag{26}$$

where, a', $b'$and c' denotes the feature of different sensor generated from the CNN models.

LSTM

The focus of LSTM in this work is to obtain the time sequence information of extracted features to achieve prediction. Once the data features of each CNN model have been merged at the concatenated layer, the generated column-wise array will be the input of the LSTM model, which can be expressed as $X = [x_{i,j}]$, where x denotes the feature value, $i = \{1, 2, \cdots d\}$, j

denotes different signals. A multi-layer LSTM uses X as input data, and each layer contains an LSTM cell (Figure 30).

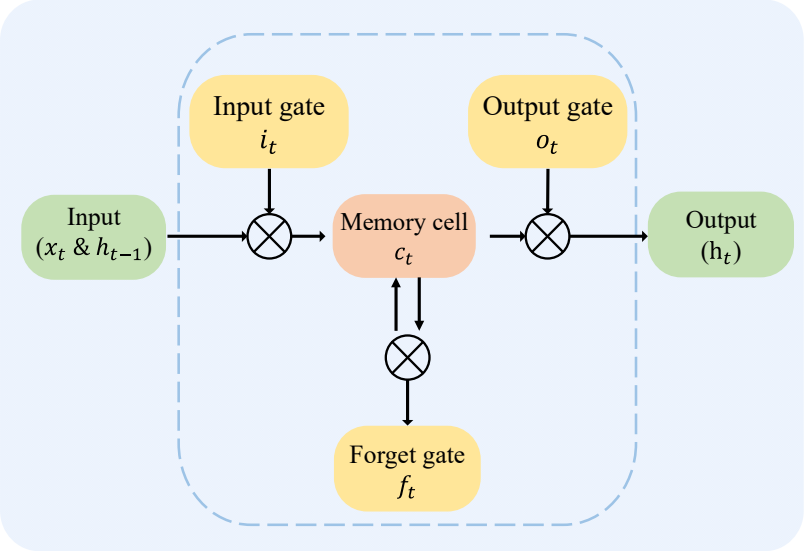


Figure 30. The LSTM cell in this work

For the current time-step t, the output h_t, which indicates the predicted flank wear value, and the memory state c_t of LSTM cell are decided by the new input x_t, the output h_{t-1} and memory state c_{t-1} of the last time step $t-1$. Based on the input array X, which contains d time-steps and j features in each time-step, the LSTM network will recursively obtain the comprehensive information from each time-step and provide corresponding predictions.

During the process, the forget gate f_t of the LSTM cell decides how many information of c_{t-1} should be forgotten according to the sigmoid activation function, and thus create the new feature data as candidate value. Then the candidate value is fed into the input gate i_t to update the memory cell state. Finally, the prediction value h_t of the LSTM cell at the time-step t can be calculated based on the updated cell state, which is from c_t and controlled by the output gate o_t. The above gate and cell can be obtained as follows:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{27}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{28}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{29}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{30}$$

$$h_t = o_t \odot \tanh c_t \tag{31}$$

where, σ is the sigmoid activation function; $\odot$ is the Hadamard product; W and U are variable weights and b is the bias.

Multiple LSTM cells are connected to form an LSTM network in time order, to predict the time sequence output (flank wear), that is $\{h_1, h_2, \cdots h_a\}$. And the number of the LSTM cell depends on the number of the feature in each signal received from the concatenation layer. The architecture of the LSTM model is shown in Figure 31.

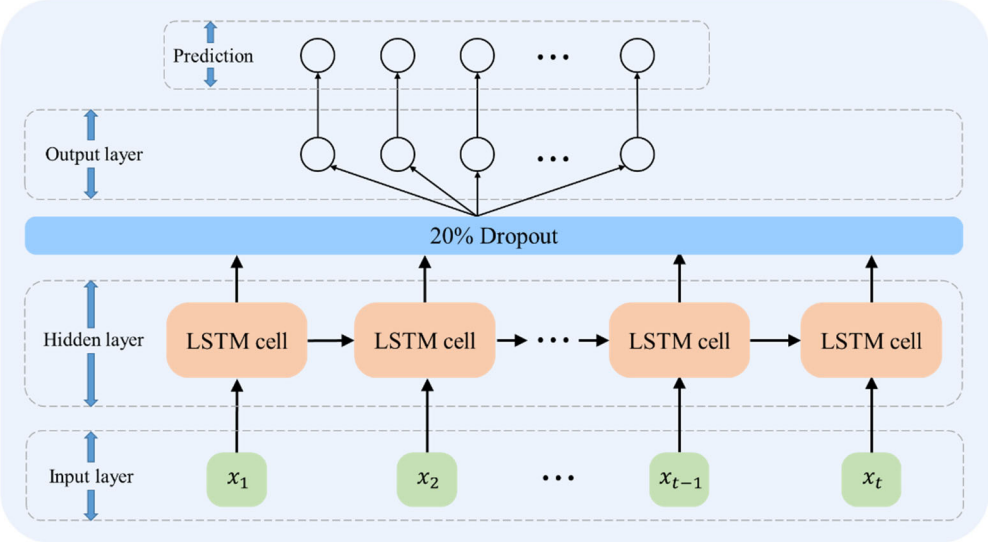


Figure 31. The LSTM architecture in this work

Since the input data is partitioned via Hurst exponent into sub-datasets in this chapter, the model training based on size-reduced sub-datasets may increase the possibility of overfitting. Therefore, inhibiting the overfitting of the proposed hybrid model is necessary to further improve the prediction performance. As the dropout is one of the most popular and efficient regularisation technologies used to prevent the overfitting, it has been employed in the LSTM model. Dropout randomly discards the hidden neuron with the setting dropout rate, and the remaining neuron is trained via backpropagation to obtain the new weight and bias (Liu et al. 2020). It can be described as:

$$S_i = \sum \sum_j w_{ij} p_j S_j \tag{32}$$

where, the S_i denotes the output of ith layer after the dropout, S_j denotes the output of the j^{th} layer (the previous layer), w denotes the weight, p denotes the dropout rate, which is set to 0.5 in this chapter to maximise the regularisation.

Furthermore, in the last output layer of the hybrid model, the linear activation function is adopted to implement the regression prediction. For the prediction evaluation of the proposed hybrid CNN-LSTM model, the mean absolute error (MAE) is used as the criteria, which quantifies the absolute error between the prediction and actual values (Bhinge et al. 2017). It can be expressed as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\tilde{y} - y| \tag{33}$$

where, n is the training sample size; $\tilde{y}$ is the prediction value; y is the actual value.

4.3 Case study and Methodology Validation

4.3.1 Experimental setup

To validate the methodologies, the dataset from the 2010 PHM Society Conference Data Challenge was adopted here (PHMSociety 2010). The dataset includes signals from cutting force sensors, vibration sensors and AE sensors that were collected in the process of dry milling on a high-speed CNC machine Roders Tech RFM760. In the experiment, a stainless-steel workpiece was machined using six 3-flute ball nose tungsten carbide cutters, and the corresponding sensor signals for the six cutting tools C1, C2, C3, C4, C5 and C6 were recorded. Every cutting tool was cut from new until significant wear, and a total of 315 cutting cycles were performed for each tool under the same machining parameters. In addition, three types of sensors are mounted on the workpiece and the machining table respectively, in terms of dynamometers, accelerometers and an AE sensor. Figure 32 shows the schematic of the experiment.

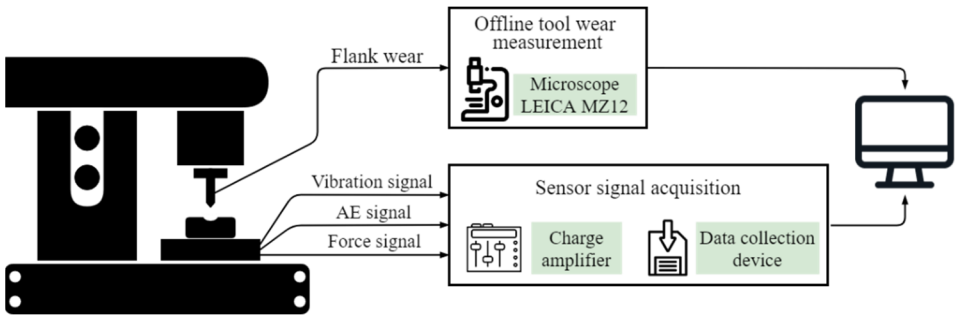


Figure 32. The schematic of the experiment

During the machining process, three Kistler quartz 3-component platform dynamometers, three Kistler piezo accelerometers and a Kistler AE sensor were used. These seven signal channels are shown in Table 23. A cutting tool processed the workpiece surface line-by-line along the X-axis with the axial depth of 0.2 mm, radial depth of 0.125 mm and cutting length of 108 mm per

cut until the entire surface was removed. The spindle speed was maintained at 10,400 RPM, and the feed rate was 1,555 mm/min. Moreover, after each cut, the cutting tools C1, C4 and C6 were placed under a LEICA MZ12 microscope to measure the flank wear of each flute, and three datasets that combined with sensor signals and flank wear of cutting tool C1, C4 and C6 were employed in this work. The dataset size of a single cutting tool is approximately 3.2 GB.

Table 23. Sensor signal of the dataset

Sensor type	Signal sources
Kistler quartz dynamometer	Force (N) in the X-axis, Force (N) in the Y-axis, Force (N) in the Z-axis
Kistler piezo accelerometers	Vibration (g) in the X-axis, Vibration (g) in the Y-axis, Vibration (g) in the Z-axis
Kistler AE sensor	AE-RMS (V)

4.3.2 Signal partition based on the Hurst exponent

Based on the datasets of the acquired sensor signal, the validation of the proposed system is implemented on a 3.60 GHz Intel (R) Core (TM) i7-7700 CPU processor (with 8.00 GB of RAM), in which, the data processing and the computation of the deep learning algorithm are executed on the Keras framework and Tensorflow, respectively.

As aforementioned, 315 cuts were performed using the cutting tools C1, C4 and C6, respectively. Taking C4 as an example for analysis, the dynamometer signals on the X-axis, vibration signal on the X-axis and the AE sensor signals for the 1^{st} cut 150^{th} cut (intermediate cut) and the last (315^{th}) cut are plotted in Figure 33. It can be observed that the amplitude of the three sensor signals grows along with the machining process, indicating that the signals exhibit an excellent association with tool wear deterioration. Thus, these sensor signals are feasible to establish the prediction.

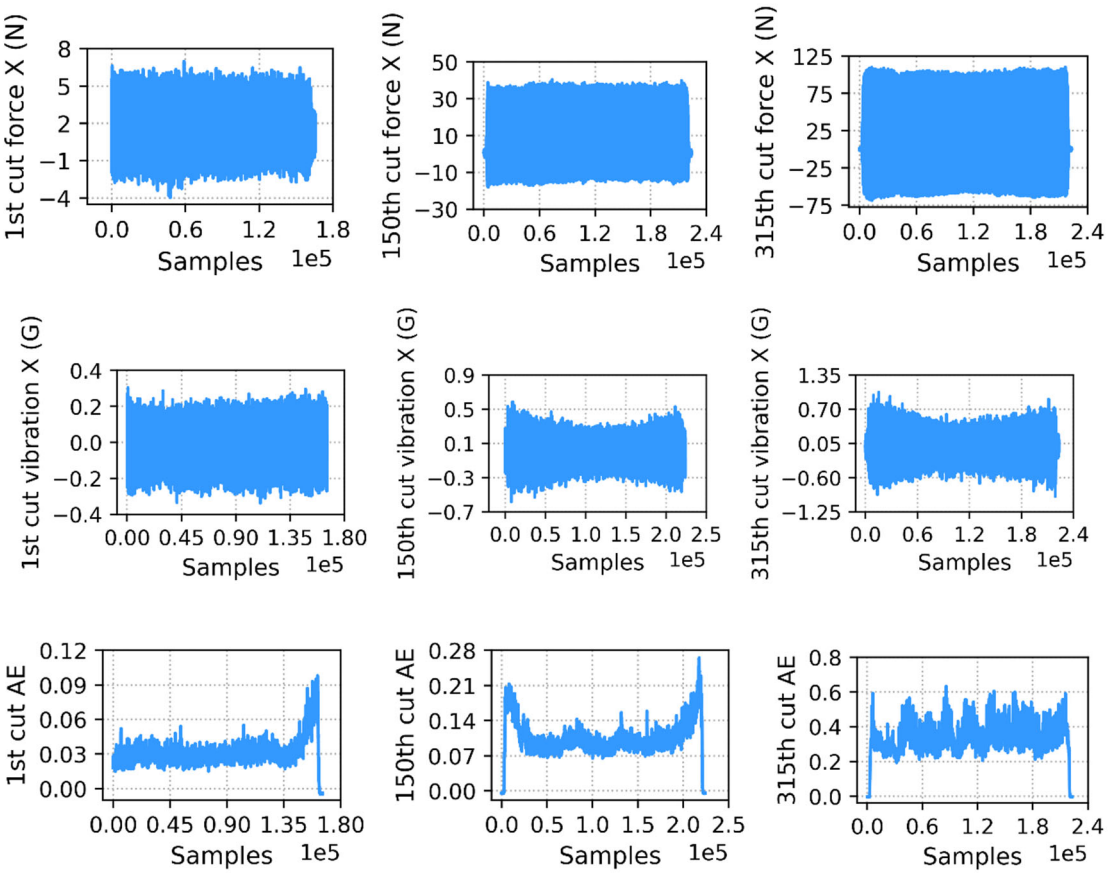


Figure 33. Sensor signals for the 1st, 150th and 315th cut by the cutting tool C4

In correspondence with each cut, the flank wear of three flutes on the cutting tool was measured during the experiment. The flank wears of the cutting tool C4 are shown in Figure 34. According to the recommendation of the ISO 8688-2 (1989), the cutting tool life criterion is commonly predetermined by the average wear value of all flutes. Therefore, the average flank wears of the cutting tools C1, C4 and C6 were used for training the prediction model presented in this chapter.

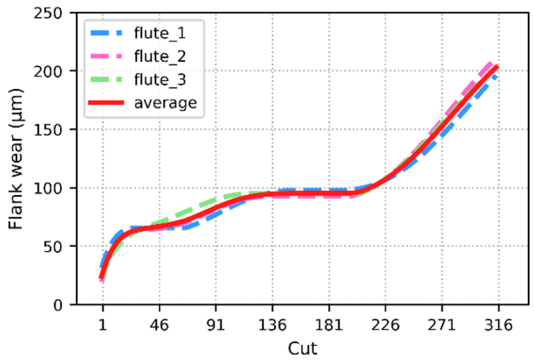


Figure 34. The flank wear of the cutting tool C4

Tool wear involves different stages. Sensor signals over each stage are usually uneven, which may cause the wear prediction inaccurate. Moreover, as discussed earlier, due to the changing features of tool wear in the different stages, it is difficult for the prediction algorithm to effectively estimate different wear trends of tool wear during these stages. In this work, the Hurst exponent is used as an index to judge the fluctuations of sensor signals, and then the signals are segmented to correspond to the stages of flank wear. For the cutting tool C4, the Hurst exponent of three types of sensor signals is calculated for the 315 cuts. The Hurst exponents of the vibration signals in the X-axis, the cutting force signals in the X-axis and the AE signal are shown in Figure 35. Furthermore, to better represent the convergence and visualisation effect of the results, a cubic curve, which is a regression analysis method that preserves data characteristics and reduces data turbulence without changing the data, was applied to fit the Hurst exponent.

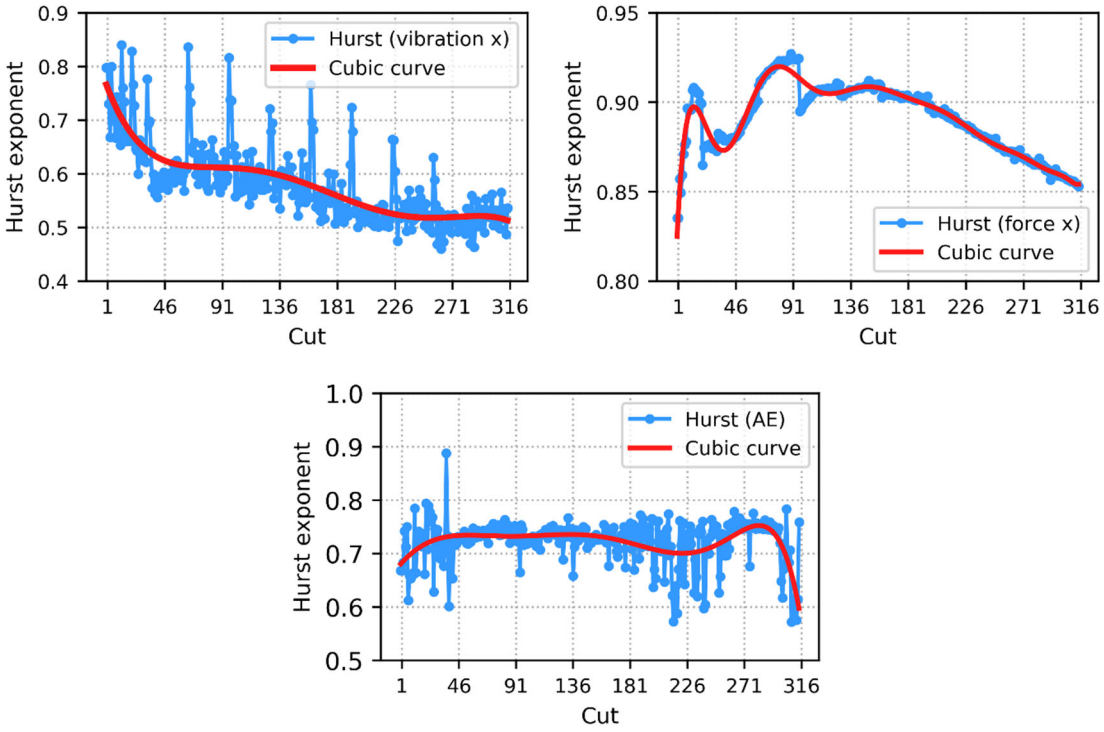


Figure 35. The Hurst exponent of the sensor signals (V, F and AE) for the cutting tool C4

From Figure 35, the Hurst exponents (the *H* values) of V, F and AE are all roughly between 0.5 and 1. It means that these sensor signals present the persistent behaviours along with the tool wear, and it is feasible to conduct prediction.

For the vibration signals, some observations are below:

- The *H* value of the cutting tool is the biggest at the beginning of machining, which is close to 0.85, implying that the signal has obvious regularity;

- As the machining progresses, the *H* values are higher than 0.6 before the 20th cut, so that the correlation between the vibration signal and the tool wear is still strong at this time. From the 20th cut to the 205th cut, the *H* values decrease to between 0.5 and 0.6, which indicates that the long-term memory of the signal is lower than its previous tool wear stage;
- After the 205th cut, the *H* values progressively approach to or even are lower than 0.5, meaning that the signals exhibit a trend of the Brownian motion. That is, the probability of negative correlation between the signal and the tool wear is increased;
- After the 315th cut, the *H* values are bound to drop below 0.5, so that the signal will completely show a negative correlation, and the cutting tool exceeds its health lifespan and the tool wear displays unpredictability.
- Overall, the fractal of the *H* values trend of the vibration signal has a significant correspondence with the different stages of tool wear, and the change of the *H* values are sensitive. Thus, the vibration signal displays the most significant potential among the three signals for signal partition.

For the signals of the cutting force and AE, here are some observations:

- Along with the machining process until the 315th cut, the *H* values both show a decreased trend, which corresponds to the tool wear. Moreover, the *H* values are both greater than 0.5, it represents these sensor signals exhibit persistent behaviours and are positively correlated with the tool wear, thereby revealing that prediction of tool wear based on the cutting force and AE signals are also feasible;
- Despite this, it should be noted, the Hurst exponent trend of these signals are not as clear as the vibration signal in performing partition on the signals; This may be interpreted as: due to the acquisition frequency of the cutting force and AE signal is lower than that of the vibration signal, the collected noise signal that generated by excessive tool wear is insufficient. Moreover, the cutting temperature gradually increases with the progress of machining, so that it will soften the material and reduce the cutting force resistance (Xu et al. 2018). And also the AE sensor is prone to be affected by the mechanical noise from the background environment (Rusinek and Borowieca 2015);
- Therefore, the signals of the three sensors are partitioned uniformly according to the Hurst exponents of the vibration signals in this work. All the signals were applied as the data sources for subsequent deep learning algorithm for the information compensation and purposes of prediction.

Accordingly, the above observations on signal changes are correlated with the stages of tool wear. For instance, for the cutting tool C4, it consists of the initial stage (1st – 19th cut), steady wear stage (20th – 204th cut), and severe wear stage (205th – 315th) as shown in Figure 36. The signal segmentations of three cutting tools are summarised in Table 24.

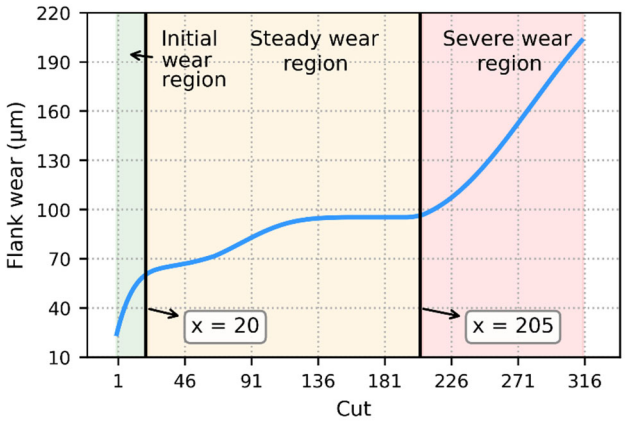


Figure 36. The stages of flank wear for the cutting tool C4

Table 24. Signal segmentation of each cutting tool

Cutting tool	Tool wear stages		
	Initial wear region	Steady wear region	Severe wear region
C1	1 to 49	50 to 139	140 to 315
C4	1 to 19	20 to 204	205 to 315
C6	1 to 14	15 to 179	180 to 315

Through the Hurst exponent, the segmentations of sensor data that correspond to different tool wear stages are identified. In this chapter, three cutting tools, i.e., C1, C4 and C6, were used. To further improve the accuracy of prediction, the segmented sensor data were pair-wisely combined as the input dataset. For example, the sensor signal and flank wear of C1 and C4 were combined as a training dataset (denoted as C1C4), and the sensor signal and flank wear of C6 was treated as the validation dataset at the same time (refer to Table 25).

4.3.3 Performance evaluation on the Hurst exponent and CNN-LSTM algorithm

To evaluate the performance of the Hurst exponent-based partition, un-segmented signals (raw signals of each sensor) were used to perform prediction on flank wear based on the designed CNN-LSTM algorithm. The prediction accuracy of each dataset in this work is obtained by its corresponding validation set. And the prediction curves for the sensor signals based on the CNN-LSTM algorithm are shown in Figure 37.

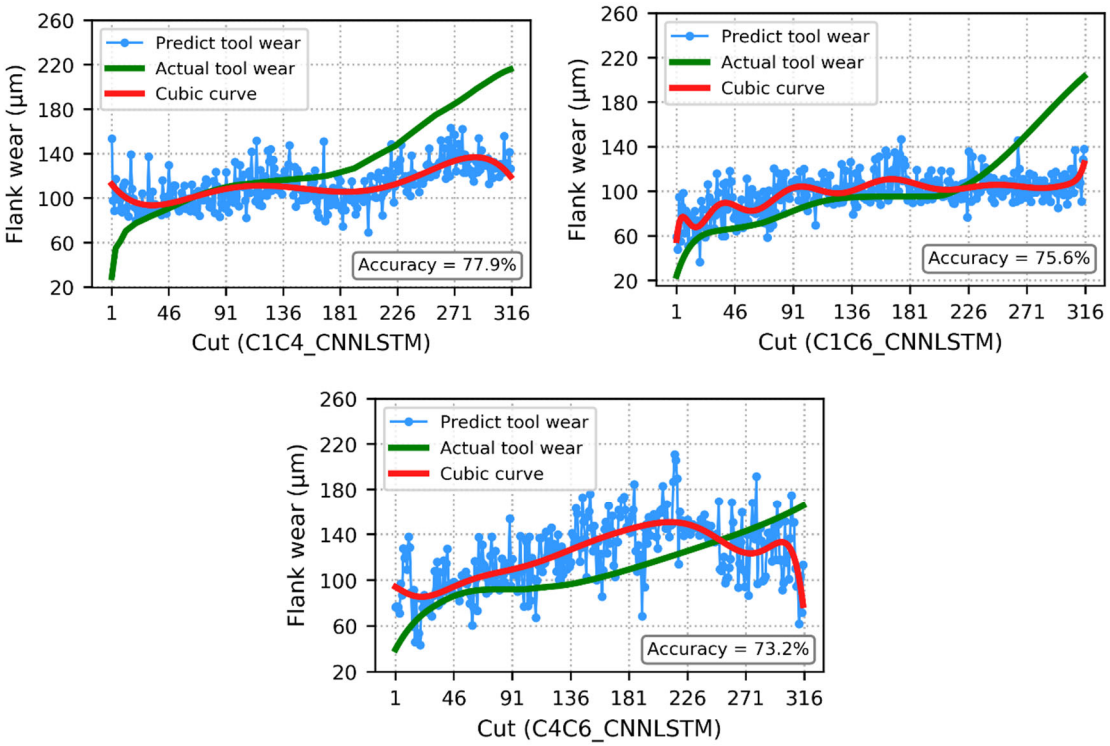
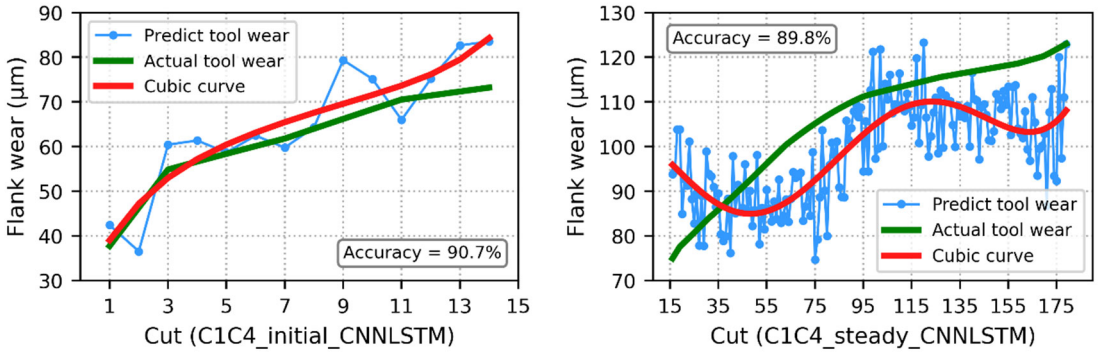


Figure 37. Prediction curves of the sensor signals based on the CNN-LSTM model

From the results, prediction accuracies are not satisfactory. It is expected to improve the prediction performance by adopting partitioned dataset according to the Hurst exponent. To do that, the sensor signals of the three cutting tools were divided into three groups according to the three stages of tool wear, and these signals were combined with corresponding flank wear values to form the new datasets. Figure 38 shows the prediction results for the signals of C1C4 based on the Hurst exponent.



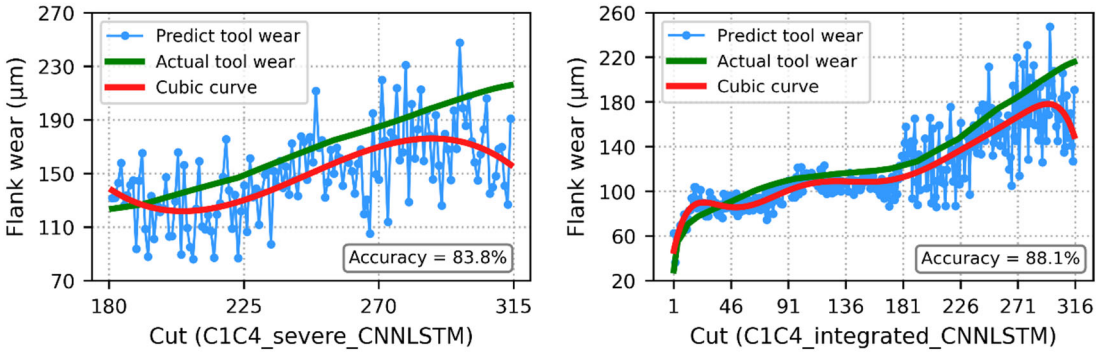


Figure 38. Prediction results based on the partitioned signals for C1C4

For C1C4, the prediction accuracies in the initial, the steady and the severe wear regions were 90.7%, 89.8% and 83.8%, respectively. The integrated prediction accuracy increased to 88.1% in comparison with 77.9% for the un-partitioned signals. The same experiments were conducted for the other datasets. As summarised in Table 25, for C1C4, C1C6, and C4C6, the accuracies using the Hurst exponent were improved by 10.1%, 10.4% and 14.5%, respectively.

Table 25. Comparisons of prediction accuracy for partitioned and un-partitioned datasets

Dataset	**Test**	**Prediction accuracy**				
		Partitioned dataset				Un-partitioned dataset
		Initial stage	Steady stage	Severe stage	Integrated	
C1C4	C6	90.7%	89.8%	83.8%	88.1%	77.9%
C1C6	C4	87.4%	88.6%	87%	86.0%	75.6%
C4C6	C1	89%	87.8%	86.3%	87.7%	73.2%

Furthermore, to evaluate the performance superiority of the developed system, the CNN-LSTM algorithm was compared with other hybrid deep learning models, which are CNN-CNN (Cheng et al. 2020), LSTM-LSTM (Choi and Lee 2018) and DNN-DNN (Zhang et al. 2019). To achieve a reasonable comparison result, the architectures and parameter of individual CNN and LSTM remained the same as the developed CNN-LSTM algorithm presented in Section 4.2. For the DNN model, it was designed to have one input layer, one output layer and three hidden layers, in which the number of neurons in the input layer is set to equal to the sample number of sensor signals, each hidden layer is set as twice the number of input signal after multiple tests, and the number of neurons in the output layer depends on the number of flank wear values. The structure of the DNN is shown in Figure 39.

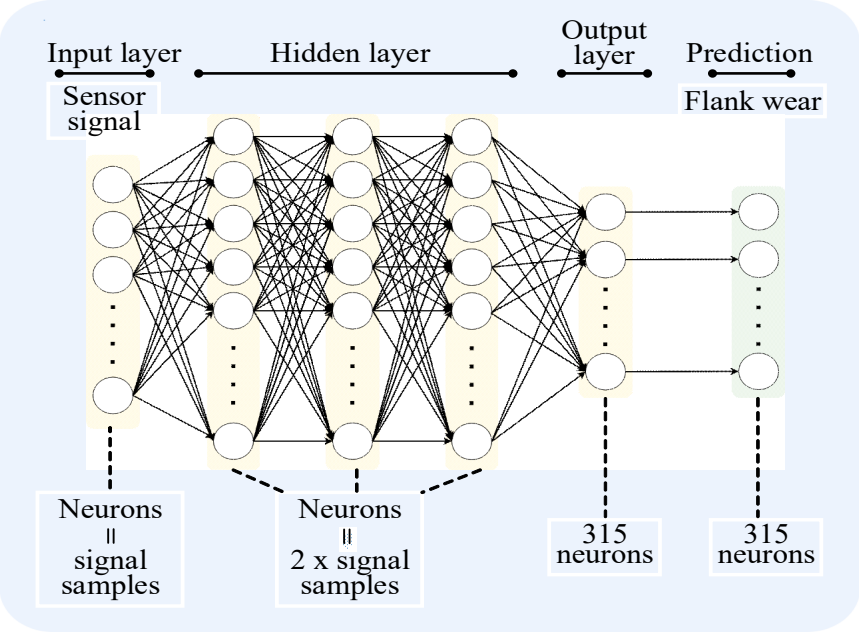


Figure 39. The architecture of the DNN model

The overall sensor signal dataset and the partitioned signal dataset were adopted to execute the prediction on the CNN-CNN, LSTM-LSTM and DNN-DNN respectively. Taking the dataset C1C4 as an example, the prediction results are presented in Figure 40 (un-partitioned signals) and Figure 41 (partitioned signals).

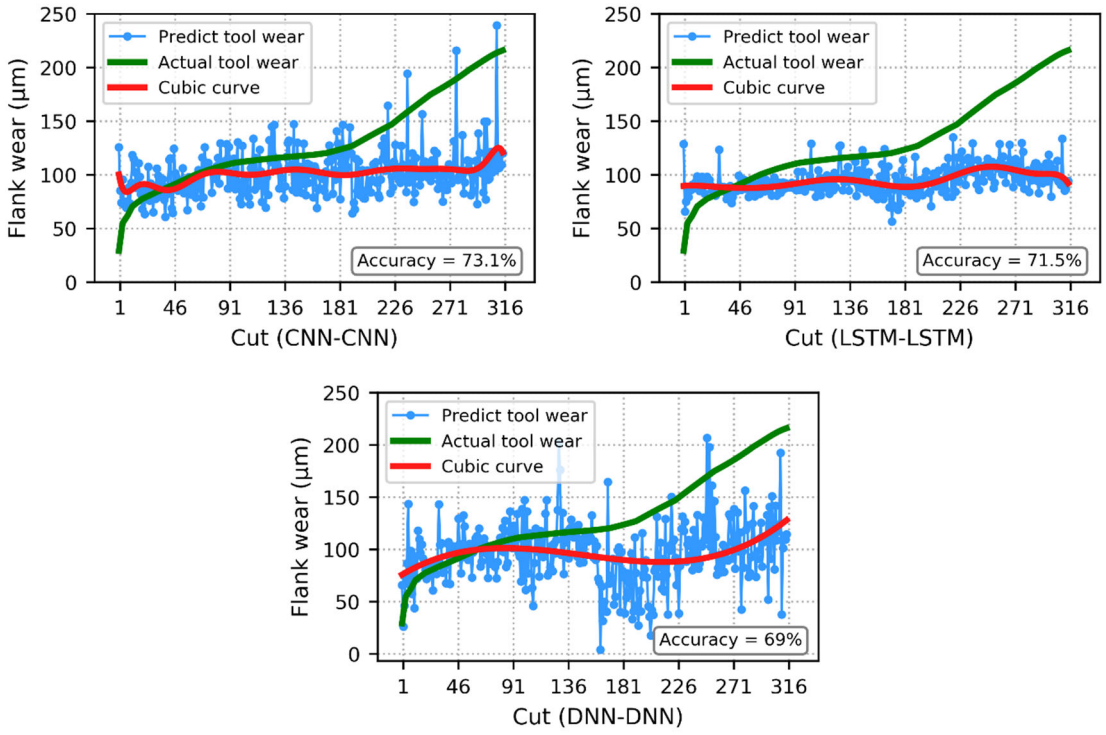


Figure 40. Prediction using the un-partitioned dataset C1C4

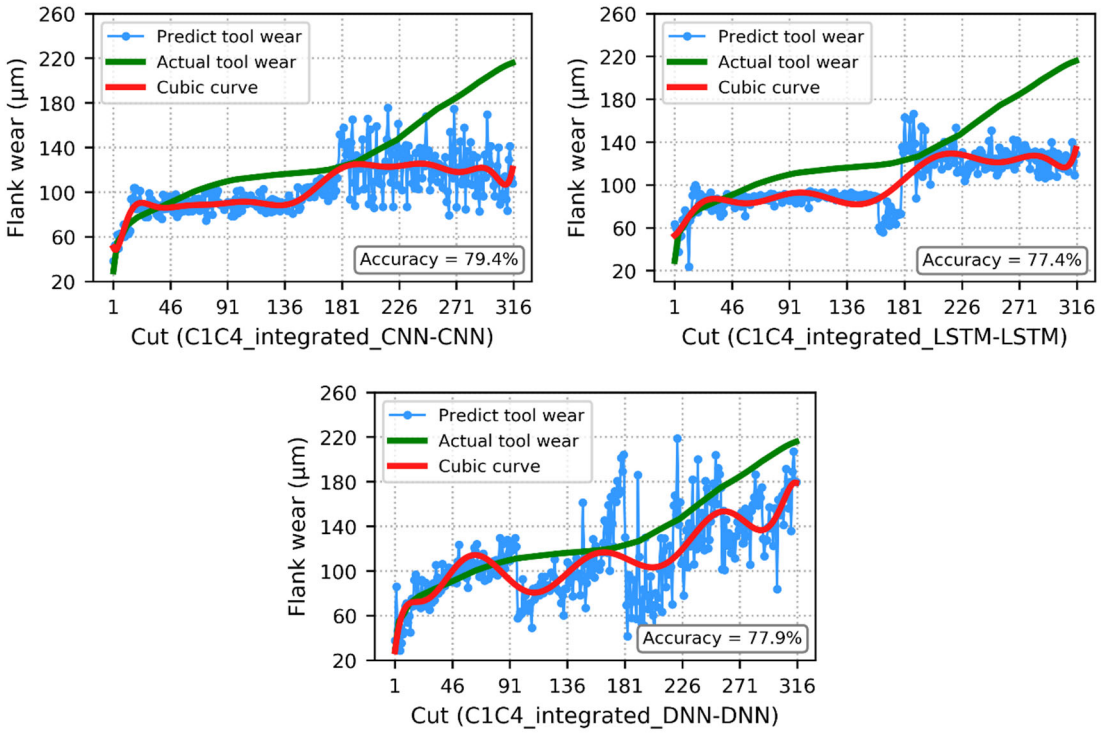


Figure 41. Prediction using the partitioned dataset C1C4

The integrated result above of the dataset C1C4 is obtained after the prediction based on the corresponding sub-dataset of the tool wear initial, steady and severe stage. Figure 42 shows several representative fitting results of the individual sub-datasets, which covers different dataset, wear stage and deep learning model, and the prediction precision of the three ensemble models on the partitioned sub-datasets and the overall datasets is summarised in Table 26.

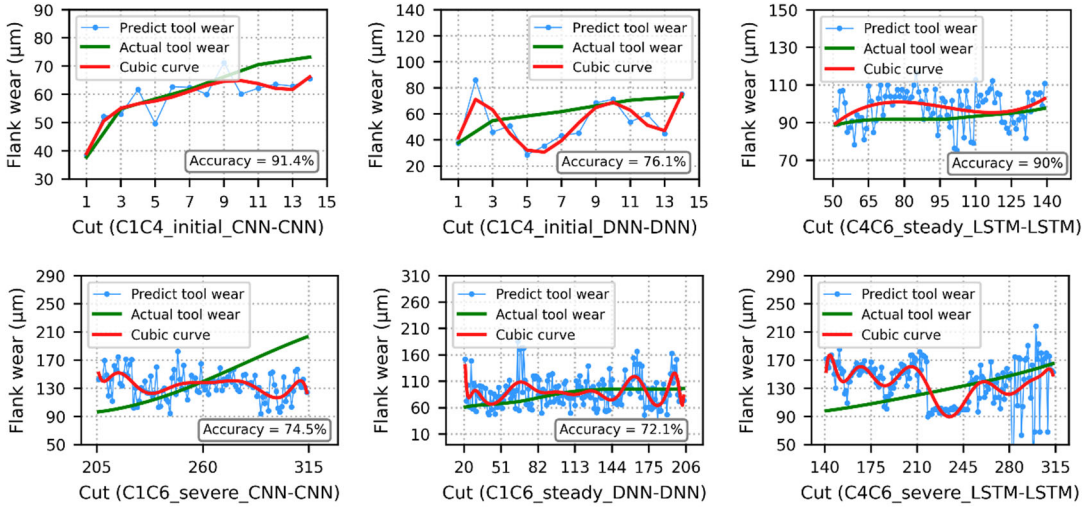


Figure 42. Curve fitting of the partitioned datasets on prediction models

Table 26. Prediction precision of CNN, LSTM and DNN model

Model	Dataset	Prediction accuracy				
		Partitioned signals				Un-partitioned signals
		Initial region	Steady region	Severe region	Integrated accuracy	
CNN-CNN	C1C4	91.4%	71.6%	81.2%	79.4%	73.1%
	C1C6	70.4%	79.5%	74.5%	76.9%	70.2%
	C4C6	81.1%	87.1%	67.7%	78.1%	70.8%
LSTM-LSTM	C1C4	80%	79.3%	74.3%	77.4%	71.5%
	C1C6	84%	78.4%	72.1%	75.5%	60.6%
	C4C6	80.7%	90%	71.4%	78%	65.4%
DNN-DNN	C1C4	76.1%	81.9%	73.4%	77.9%	69%
	C1C6	76.8%	72.1%	70.3%	71.6%	69.5%
	C4C6	72.7%	83.6%	70.7%	74.5%	63.9%

By comparing the results of Table 25 and Table 26, it is evident that, even if the un-partitioned datasets were adopted as the input, the prediction accuracy of the CNN-LSTM algorithm was better than those of other deep learning algorithms, which the average accuracy of the CNN-LSTM algorithm is 75.6%, while those of the CNN-CNN, LSTM-LSTM and DNN-DNN algorithms were 71.3%, 65.8% and 67.4%, respectively. It benefits from the proposed hybrid model that integrates the advantages of the automatic feature extraction of CNN and the time-series data sequence learning of LSTM. Moreover, from the perspective of employing the partitioned dataset, the accuracies of the four algorithms were all improved to some extents. The CNN-LSTM algorithm achieved the best average performance of 87.3%, followed by the CNN-CNN algorithm of 78.1%. Such a result proves again that, the proposed data partition strategy of the Hurst exponent is effective, which the segmented sensor data is beneficial for optimising the prediction.

Computing efficiency is also a vital factor to investigate. In this chapter, the cumulative calculations of the computing time cost were carried out for the deep learning algorithms. The time usage of each prediction model to compute un-partitioned datasets and partitioned datasets are shown in Figure 43.

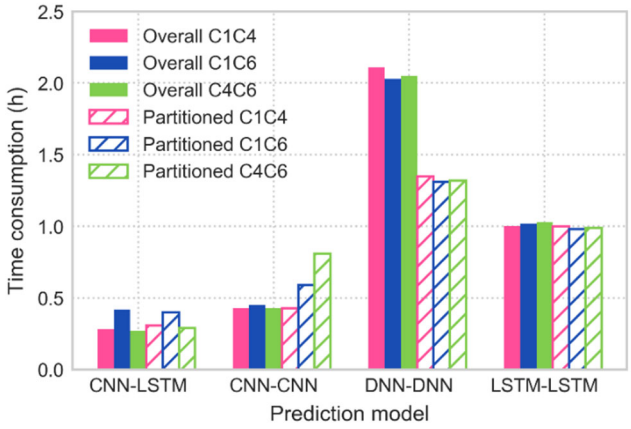


Figure 43. Time consumption of the different prediction algorithms

From the aspect of model training time consumption, it can be observed that whether the un-partitioned or the partitioned data as the input for training, the proposed CNN-LSTM model consumes the minimum computing time among all four algorithms. The average time for the CNN-LSTM algorithm was approximate 0.33h, the average time cost of the CNN-CNN model was the closest to the CNN-LSTM algorithm, about 0.52h. The DNN-DNN and LSTM-LSTM algorithms consume 80% and 67% more time than the CNN-LSTM algorithm, respectively. In view of the CNN-LSTM model being outstanding in the prediction precision and efficiency, it can be considered a powerful and promising deep learning scheme for the cutting tool RUL regression prediction.

4.3.4 Comparison of the Hurst exponent with other methods

Furthermore, to effectively investigate the adaptability of the Hurst exponent-based signal partition method, the performance comparison between the proposed method and other prevalent signal processing approaches were executed. These approaches are domain-feature extraction (Huang et al. 2020) and PCA dimensionality reduction (Lee et al. 2020).

The signal processing of the prediction task is responsible for eliminating unwanted data from massive sensor signals, for improving the efficiency and accuracy of the prediction model with lower volume and valuable data. For the domain-feature extraction, the features of the signals adopted in this chapter were extracted under the time-domain, frequency-domain and time-frequency-domain. Under the time-domain, statistical features are usually extracted to reflect the change of signal properties over time. And, as another indispensable feature extraction method, the frequency-domain signal is able to present the signal change rate and provide the spectral feature for hidden information revealing (Herff and Krusienski 2018). The fast Fourier transform (FFT) is the most common frequency-domain transformation. It is employed to convert the sensor

signals of each cut cycle in this work. The FFT of the X-axis vibration signal of the 1st cut of the cutting tool C1 is shown in Figure 44. Moreover, facing the nonlinear sensor signal, the time-frequency-domain signal outperforms the methods to process signals under the time-domain and frequency-domain by capturing useful features, since it is conducive to explore the feature in transient and localized components. In this chapter, the continuous wavelet transform (CWT) is applied to transform each sensor signal to the time-frequency-domain, which is a powerful method to represent the sensor signal into a two-dimensional plane (Lee et al. 2017). The CWT result of vibration signals of the 1st cut for C1 in the X-axis is shown in Figure 45.

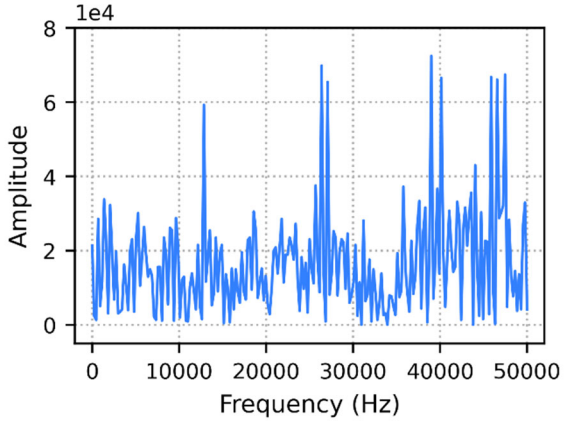


Figure 44. FFT spectrum of the X-axis vibration signal of 1st cut of the cutting tool C1

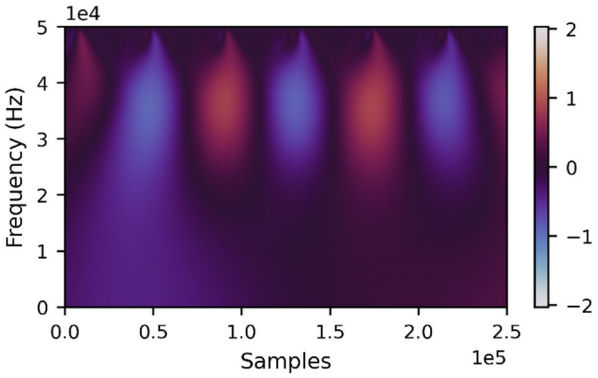


Figure 45. The CWT plot of the X-axis vibration signal of the 1st cutting for the cutting tool C1

After the conversion of the frequency-domain and time-frequency-domain for sensor signals of the 315 cuts of each cutting tool, including the time-domain, several widely employed features are extracted in every domain. It is noteworthy that the feature extraction in the time-frequency-domain is based on a different scale, which is divided according to the sampling rate of the adopted dataset. Table 27 summarises the extracted features of the three domains in this chapter.

Table 27. Extracted feature from the different domain

Domain			
Time-domain	**Frequency-domain**	**Time-frequency-domain**	
Mean		0-10000 Hz	
Standard deviation	Frequency centre	10000-20000 Hz	Mean
Skewness	Median frequency	20000-30000 Hz	Standard deviation
Kurtosis	Root variance frequency	30000-40000 Hz	Variance
Maximum		40000-50000 Hz	
Peak to peak			

The validation dataset in this chapter contains 7 sensor signals for the 315 cut cycles of 3 cutting tools. Finally, there are 158,760 features obtained from the raw sensor signal by the above feature extraction mentioned. Then, the generated features of each cutting tool are gathered as a new input dataset, and they are adopted to perform the flank wear prediction on the CNN-LSTM algorithm, the fitting result is shown in Figure 46.

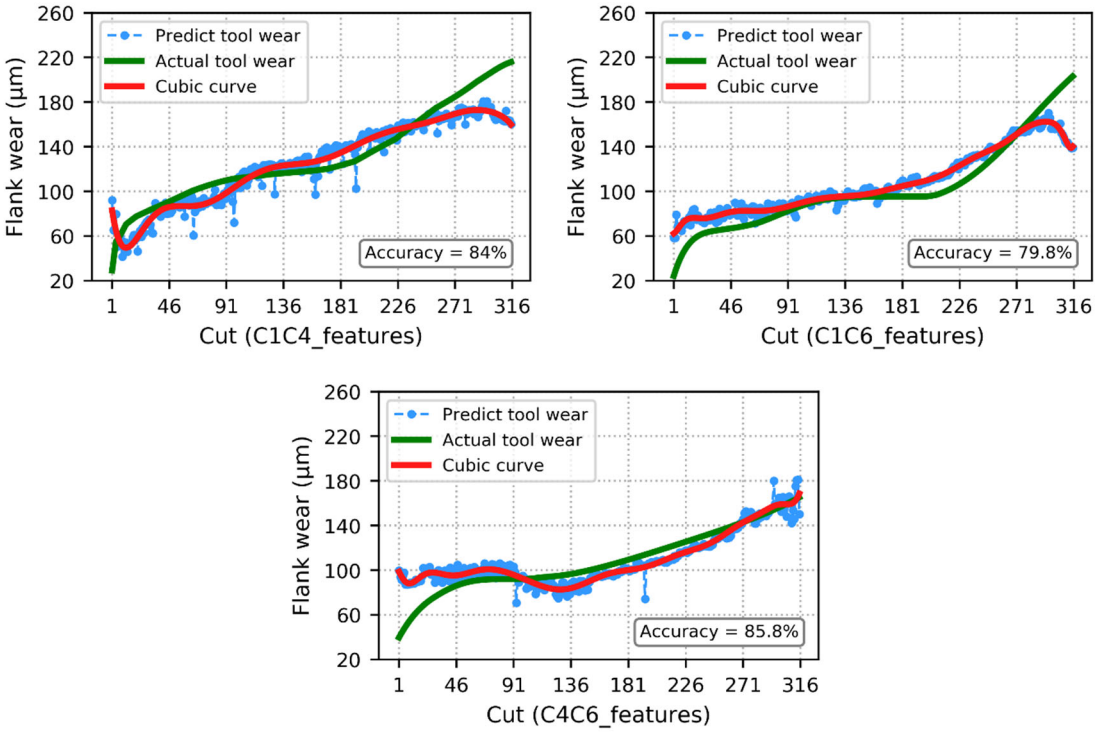


Figure 46. Fitting results of the domain feature input by the CNN-LSTM algorithm

Besides the domain-feature extraction, the signal processing based on the PCA is a prevailing approach for dimensionality reduction of the multi-sensor signal. It is capable of compressing the

signal dimension without discarding primary information by mapping the raw sensor signal into a feature space with representative components (Li et al. 2017). For the dimensionality reduction of the sensor signal by PCA, the number of components is necessary to be determined initially via the cumulative sum of explained variances, which provides the proportion of the retain information of the different number of components. Taking cutting tool C1 as an example, the cumulative sum of explained variances of the sensor signal in the 1st cut is shown in Figure 47.

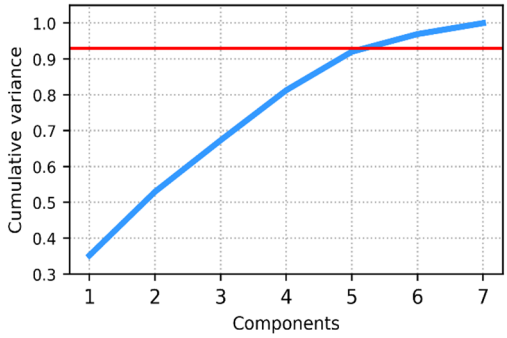


Figure 47. The PCA cumulative variance plot for the 1st cutting of the cutting tool C1

From the above figure, it is found that five components generated by the PCA could remain over 90% original sensor signal information, which is the same outcome of the other two cutting tools. Thus, these five components are adopted to replace the raw sensor signals of every cutting tool, each component includes the same number of data as the original sensor signal. Then, the flank wear prediction based on the five components was executed by the CNN-LSTM algorithm. And the input strategy is pair-wise concatenated as well. Figure 48 shows the prediction result.

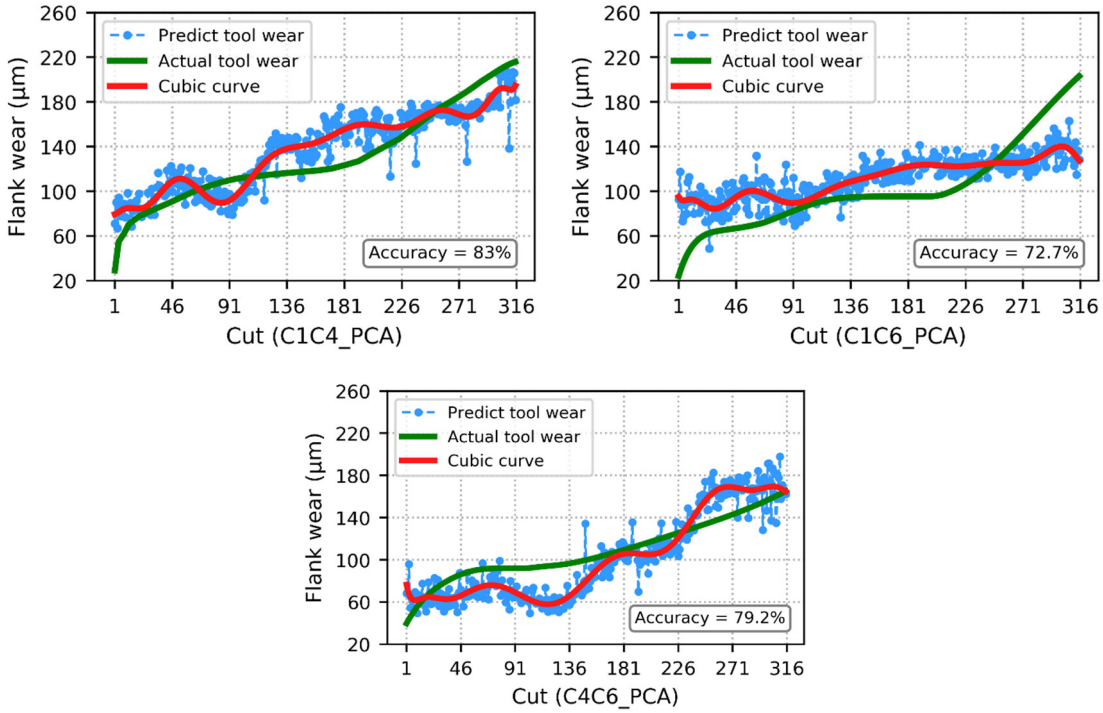


Figure 48. The fitting results of the PCA component input on the CNN-LSTM algorithm

Based on the described approach of the domain feature extraction and the PCA dimensionality reduction, the corresponding flank wear prediction result on the hybrid CNN-LSTM model are detailed in Table 28. Figure 49 shows the total processing time consumption of the sensor signals of the three datasets using these two methods and the proposed Hurst exponent method, as well as the cumulative time for their subsequent predictions.

Table 28. Prediction results of the different data processing methods

Data processing method	Input dataset	Prediction accuracy	Average accuracy
Domain feature extraction	C1C4	84%	83.2%
	C1C6	79.8%	
	C4C6	85.8%	
PCA dimensionality reduction	C1C4	83%	78.3%
	C1C6	72.7%	
	C4C6	79.2%	

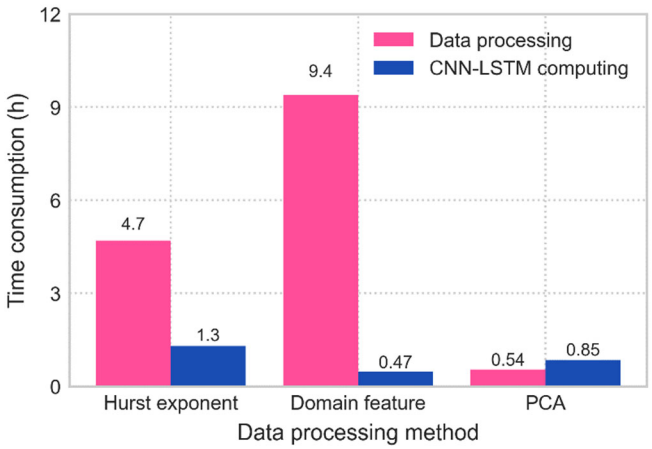


Figure 49. Time consumption of different data processing methods

By comparing the result of Table 25 and Table 28, it is apparent that the prediction accuracy for tool wear based on the dataset partitioned through the proposed Hurst exponent method was higher than applying the domain feature extraction and the PCA dimensionality reduction. Separately, the feature extraction in the three domains achieved the data volume decreasing to a great extent. Due to the elimination of redundant signals, its precision was closer to the Hurst exponent-based partition method and the model calculation time was also significantly compressed. However, the domain conversion and feature extraction consumed a large amount of time, which were 50% and 94% higher than those of the Hurst exponent method and the PCA dimensionality reduction, respectively. It is an obvious obstacle to practical applications. On the

other hand, PCA showed the benefit of the computational time-cost saving, by condensing the raw sensor signal into five new representative components, an approximate 30% dimensionality reduction was implemented. Nevertheless, the prediction accuracy is sacrificed to reach this dimensionality reduction, and it is foreseen that the optimal components of the PCA will appear more deviation along with the increasing of the signal volume. According to these assessments, the introduced Hurst exponent-based signal partition displays a more robust ability than the other two prevalent data processing approaches. Its effectiveness for the flank wear prediction could be illustrated as that the Hurst exponent partition method simplifies the system frame with its non-complex computation, and helps accomplish the satisfying performance of the deep learning algorithm on the premise of not discarding the signal information.

4.3.5 RUL prediction

After determining the superiority of the Hurst exponent partition method for the flank wear prediction, as the most ordinarily used tool life criteria, the flank wear further defines the cutting tool RUL based on the above prediction results. In general, the threshold of tool wear and tool life is dependent on the application scenario and demand. Based on the dataset employed in this chapter, the life of each cutting tool is defined as 315 cycles, and the remaining processable cycle is used to indicate the RUL. Thereby, the linear inverse relationship between the processed cycle and the remaining processing cycle can be established intuitively, similar to the work of (Mao et al. 2018; Li and Liu 2019). Besides, the degradation of the actual remaining processable cycle is caused by the tool wear (measured flank wear). Therefore, the predicted remaining processable cycle can be further calculated with the predicted flank wear as follows:

$$\mathrm{RUL}_p{}^i = \frac{{w_p}^i \cdot \mathrm{RUL}_a{}^i}{{w_a}^i} \tag{34}$$

where, RUL_a denotes the assigned actual remaining processable cycle, which is $\{315, 314, \cdots, 1\}$. w_p denotes the predicted flank wear, w_a denotes the actual measured flank wear, i denotes the number of the cuts.

A polynomial regression fitting model is then constructed for the RUL prediction of the cutting tools based on the sample data obtained by Equation 34 and the predicted flank wear. The regression function can be described as:

$$F(w) = \sum_{j=0}^{k} b_j \cdot w^k \tag{35}$$

where, F denotes the RUL of a cutting tool, w denotes the flank wear, b_j denotes the regression coefficients.

Figure 50 shows the comparison of the RUL prediction results of the partitioned and the un-partitioned datasets. And the RUL polynomial regression result of the partitioned datasets is depicted in Figure 51.

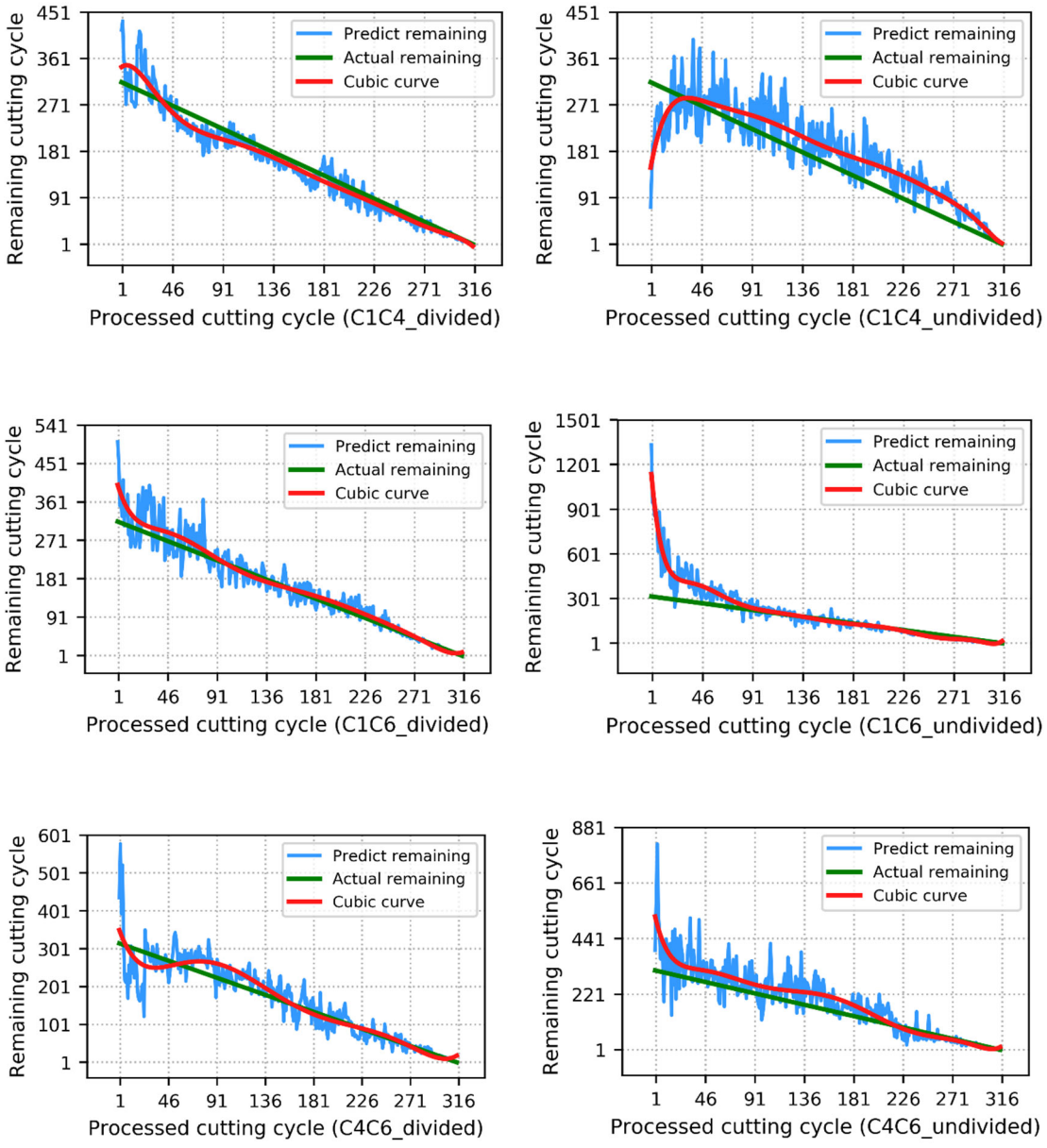


Figure 50. Predicted RUL and actual RUL of 3 cutting tools using partitioned and un-partitioned datasets

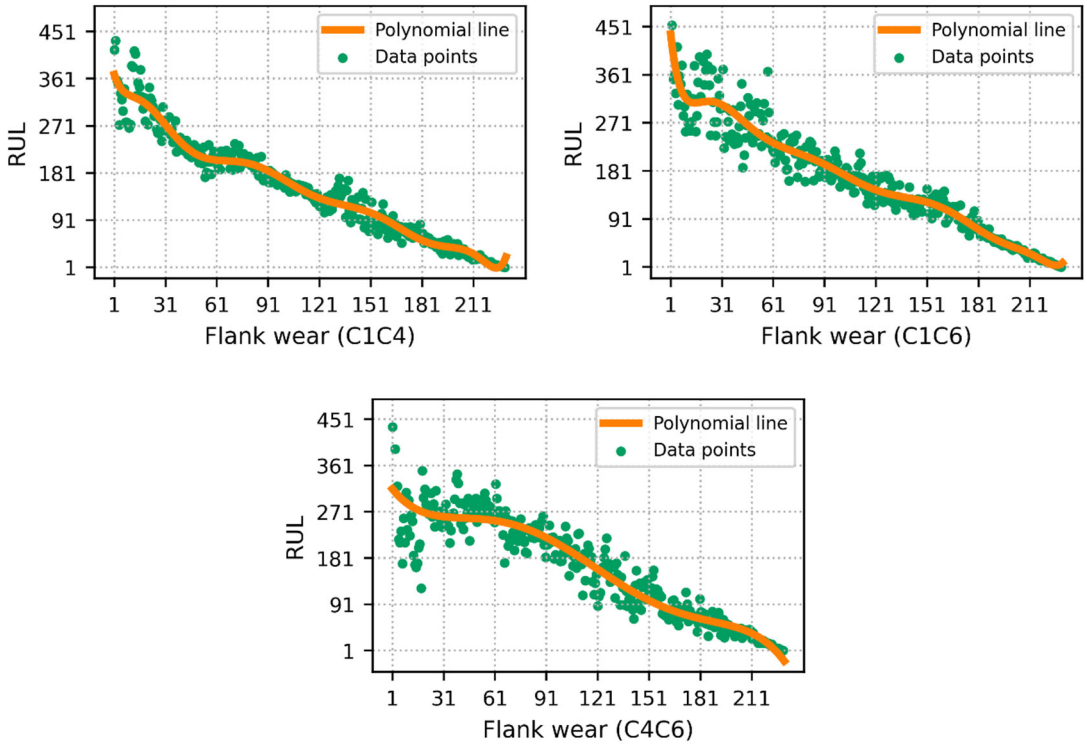


Figure 51. Polynomial regression of the RUL of 3 cutting tools using partitioned datasets

It can be observed from Figure 50 that the predicted value of the RUL varies round the actual value, the overall trend persists unchanged.

Throughout the entire machining process, the predicted value always shows fluctuations near the actual value, the relative error is large at the beginning of the machining, and the predicted value of the tool life closely resembles the actual value during the subsequent machining.

With the tool wear deteriorates, the remaining machining cycle gets easier to estimate. From the perspective of the segmented dataset C1C4, it can be noticed that between the 1st cut and the 91st cut, the predicted cutting tool RUL fluctuates greatly and has certain errors. However, as the cutting continues until reaching the end of its life cycle (the 315th cutting cycles), the predicted value begins to approach the real value, and the prediction accuracy is guaranteed. In contrast, the prediction error of the un-partitioned dataset is still huge, in specific, the RUL of the cutting tool is underestimated in the interval from the 1st cut to the 20th cut, which may lead to early replacement of the cutting tool, and results in undesired waste.

After the 20th cut, the remaining life of the tool has been overestimated, which will cause the poor quality of the workpiece surface. Moreover, the prediction error of the undivided C1C6

dataset is more obvious that reaches the highest of 74%, and the undivided C4C6 dataset shows the overestimation in a long duration.

Additionally, the polynomial regression in Figure 51 displays that the model effectively fits the flank wear and the RUL, it can prove that the model has an excellent performance in RUL prediction. The accuracy of each prediction is assessed via the coefficient of determination (R^2), which evaluates the extent of the model interpret and predict the result, the R^2 value of 0.893, 0.898 and 0.856 are achieved by dataset C1C4, C1C6 and C4C6, respectively.

In the view of all results displayed above, the prediction accuracy of the dataset that partitioned by Hurst exponent presents significant improvement than that of the unsegmented dataset, and it illustrates the cutting tool RUL prediction system proposed in this chapter has the potential and superior performance.

4.4 Summary

For the improvement of the prediction efficiency of the cutting tool RUL under the manufacturing environment of big data. This chapter presents a new methodology integrating the Hurst exponent and the hybrid CNN-LSTM algorithm in a systematic means. In this work, a Hurst exponent-based scheme is developed to partition the signals of vibration, cutting force and AE collected along the lifecycles of a set of cutting tools. A hybrid CNN-LSTM algorithm is then designed to combine feature extraction, fusion and regression based on the multi-sourced and segmented signals to predict the flank wear and RUL of cutting tools. A case study with a set of complex sensor signals was used to validate the developed methodology. To depict the performance of the proposed system, a set of benchmarks with comparative algorithms, including CNN, LSTM, DNN, was conducted under the conditions of partitioned and un-partitioned signals. Additionally, the proposed CNN-LSTM algorithm has also executed the prediction based on the dataset that processed by the feature extraction (in the time, frequency and time-frequency-domain) and PCA dimensionally reduction. Results showed that, based on the case study in this chapter, the prediction accuracy of the proposed method reaches 87.3%, which are significantly superior to other benchmarking algorithms. Analyses of the results and observations were given in detail.

Chapter 5. EC enabled Multi-sensor Tool Prognosis IoT System

5.1 Introduction

In order to improve the real-time processing ability and the data security along with reducing data bandwidth of cloud-enabled IoT system, a comprehensive real-time tool condition prognosis system based on the EC has been proposed in this chapter. This system interrelates with two modules, namely hardware and software. For hardware, an IoT monitoring platform is primarily built based on WSN, which using cost-efficient and highly openness devices in terms of sensors and microcontroller for the multiple sensor signal collection, and the platform has robust scalability and versatility to integrate different sensors easily. Moreover, in addition to communicating with the signal acquisition platform and delivery data to the CC centre, the EC configured in the proposed IoT platform is also responsible for providing stable and affordable-device-applicable methods for the signal data processing, regards to the signal denoising and the signal compression that converts the multi-sensor data into the 2D format. Meanwhile, the EC end executes the real-time tool wear identification with a hybrid 2D CNN-RF deep learning model by adopting the decision-level fusion strategy. Furthermore, in light of the greyscale image converted from the multi-sensor signals, the stacked CNN model deployed at the cloud server cooperates with LSTM, to perform tool RUL prediction using historical and future data. The proposed system provides the desired performance in line with the big data environment and it has been validated by a real machining experiment.

The remaining part of the chapter is arranged as follows. In section 5.2, the framework of the sensor acquisition platform and working principles of sensors will be explained. Section 5.3 shows the design and procedure of machining experiment to validate the proposed monitoring platform, followed by the analysis of collected data in Section 5.4. Then, section 5.5 presents the IoT monitoring system development, in terms of EC enabled system, signal processing technologies and the prognosis models. Section 5.6 is the experimental validation of the proposed EC based IoT prognosis system, and the result analysis. The summary is in Section 5.7.

5.2 The multi-sensor signal acquisition platform

For the purpose of meeting the needs of the industrial for low-cost, flexible and easily implementable monitoring system, a multi-sensor equip-able signal collection platform has been

developed based on the Arduino board, which is an open-source micro-control platform cost only US $25, and can be combined with a variety of sensors or devices (Lockridge et al. 2016).

The Arduino board that is selected as the microcontroller for the signal acquisition platform is not only because its reasonable price, also it has a simple programming environment (C/C++ language integrated development environment), which offers the high degree of freedom and expandability. Moreover, the standardized port of the Arduino board has laid the foundation for its optimized development, and the large-scale user communities and applications in various fields also benefit from the open-source features of Arduino (Fatehnia et al. 2016). In the proposed platform, a typical board Uno of Arduino is adopted, it equips with various types of connection ports, including digital input and output, UART TTL (5V) serial communication, PWM output and analogue input (Karami et al. 2018). Moreover, the 5 V and 3.3 V output can be provided by the board for the necessary power supply voltage of sensors. In addition, the Arduino Uno does not require the specific operating system and is compatible with many shields (extension board) to implement corresponding extra functions, like Wi-Fi shield.

Furthermore, to enhance the flexibility and reduce the implementation cost, a wireless communication solution is considered in this platform. Besides, a local database has been developed to host and store the multiple sensor data, and the Apache Spark framework is adopted to organize acquired numerous data and to support further data processing and analysis. For the ability evaluation of the platform, three types of sensors have been selected to monitor vibration, cutting force and energy consumption profiles of a CNC milling machine. The framework of the monitoring platform is depicted in Figure 52.

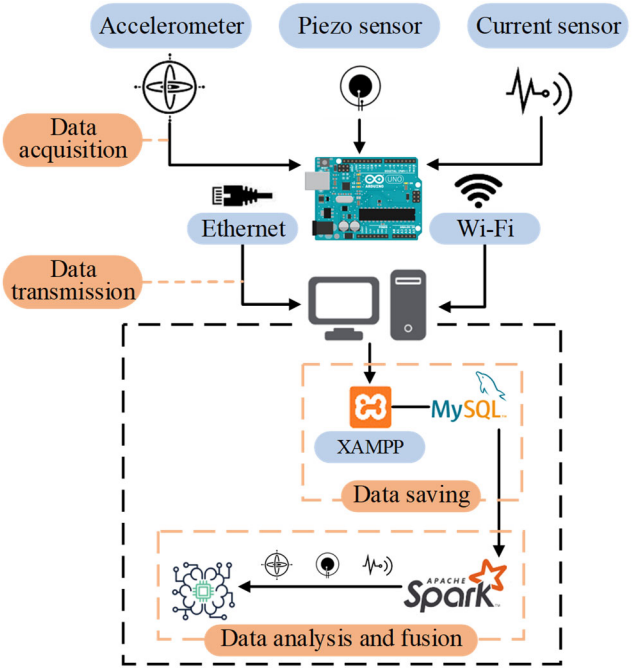


Figure 52. The monitoring platform framework

5.2.1 Vibration sensor

The friction and contact between the cutting tool and the workpiece lead to different tool conditions such as tool wear, and the most direct manifestation of this interaction is vibration. Therefore, the measurement of the vibration signal has become an essential task for TCM.

The vibration sensor collects the vibration signal of the cutting tool, and converts the non-electrical vibration signal into an electrical signal through physical action (Ostasevicius et al. 2015). According to the different principles of vibration sensors, it can be divided into six types. Table 29 shows their advantages and disadvantages (Chaurasiya 2012).

Table 29. The advantages and disadvantages of different vibration sensor types

	Advantages	**Disadvantages**
inductive sensor	Simple structure, Non-contact measurement,	Low sensitivity, Limited by the detected metal material,
Piezoelectric Sensors	Wide frequency bandwidth, Simple structure, Light weight,	High resonant frequency, High output impedance, Weak output signal, Easy to be interfered by environment,
Magnetic Sensors	Large output signal, Strong anti-interference ability,	Complex structure, Expensive,
Capacitive Sensor	High resolution, Wide measurement range, High precision, Short response time, Suitable for online, Non-contact measurement,	High output impedance, Affected by electromagnetic fields,
optical fiber sensor	Light weight, small size, Short response time,	Complex structure, expensive, Requires precise installation technology,
Photoelectric Sensor	High resolution, Short response time, Non-contact measurement,	Limited measurement range, Susceptible to environmental, Difficult for industrial applications,

The comparison of the above table clearly shows the vibration sensors of different principles have their own characteristics, the capacitive vibration sensor is a more feasible option for the affordable signal acquisition platform application with its obvious advantages. Therefore, the developed platform adopted the capacitive three-axis accelerometer MMA7361 (specification shown in Table 30) for vibration signal acquisition. This sensor is a breakthrough for three-axis

accelerometer with low power and reasonable price, which is more compatible with the Arduino Uno board to provide a stable measurement.

In practical, the acceleration measurement of MMA7361 is based on the principle of capacitance, which makes it simple and requires only a single chip to measure the acceleration of triaxle, the acceleration fluctuation reflected via the displacement change of the central mass of the chip between the fixed beams of the capacitance. The formula for the MMA7361 to convert the ADC input value to the acceleration is depicted as follows (Open Energy Monitor 2020):

$$Acc = \frac{\frac{adc\_input \times ref\_v}{1023} - v_0}{sen.} \tag{36}$$

where, Acc denotes the acceleration of each axial, adc_input denotes the raw value from the accelerometer, ref_v denotes the voltage of sensor supply, v_o denotes the voltage at 0 acceleration, $sen.$ denotes the sensitivity of the accelerometer.

Table 30. Specification of the 3-axis accelerometer MMA7361

Bandwidth Response	400 Hz
Selectable Sensitivity	±1.5g, ±6g
Sensitivity	800 mV/g @ 1.5g
Operating voltage	5 / 3.3 Volts
Current Consumption:	400 μA

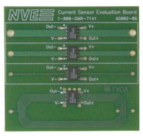
Furthermore, the calibration of the accelerometer may be necessary before the monitoring started. Based on the MMA7361, the calibration follows the principle: the output values of the X- and Y-axis should be 0 because the accelerometer is stationary; the output value of the Z-axis is 100 that means the acceleration equal to the gravity acceleration (as the unit of the output value is $g \times 10^{-2}$, and g is the acceleration of gravity).

5.2.2 Current sensor

In many monitoring systems, the motor power is considered to be one of the most suitable signals for workshop applications (Ziegler et al. 2009), its principle is not complicated and the monitoring process does not require the machine shutdown. Thus, as the signal source of the power, the importance of current sensors is reflected in condition monitoring. There are three types of current measurement technologies based on different fundamental principles. For example, shunt and PCB Trace Resistance are developed from Ohm's Law, Faraday's Law of

induction is the fundamental of current transformer and Rogowski coil, and Hall Effect sensor is based on the magnetic field. The advantages and disadvantages of commonly used current sensors corresponding to these technologies are summarised in Table 31 (Costa et al. 2015).

Table 31. Pros and cons of different current sensors

Sensor	Sensor type	Advantage	Disadvantage
YHDC SCT013 	Current transformer	• High accuracy with wide dynamic range and high frequencies • Low cost • Non-contact sensor • Able to measure high current, with low power consumption	• Measure AC only
TIDA-03040 	Shunt	• Low resistance • Low inductance winding	• Suitable for the low current • Need connect to the circuit under test • Thermal drift
AG003-01E 	PCB Trace Resistance	• Low cost • Space saving • Low resistance	• Need connect to the circuit under test
553-2177-ND 	Rogowski coil	• Low cost • Non-contact sensor	• Output signal weak, amplifier is needed • Sensitivity to magnetic fields
US5881 	Hall-effect sensors	• Non-contact sensor	• Sensitivity to magnetic fields • Signal conditioning needed

Through the comparison above, it could be noticed that shunt and PCB trace resistance not able to apply for the TCM system, because they demand direct contact with the circuit under test. Moreover, Rogowski coil and Hall-effect sensors are sensitive to surrounding magnetic fields, which is not adaptable in the industrial environment and the collection stability could not guarantee. All things considered, the current transformer is the best choice of the current signal collection for the proposed signal acquisition platform.

As a typical current transformer, YHDC SCT013 (specification in Table 32) is a widely employed current measurement sensor, which is based on the principle of electromagnetic induction, the current passes through the core magnetic field of YHDC sensor, cutting the magnetic induction line, and then current in the winding of YHDC have been generated. Due to the YHDC current sensor describes the average strength of the current, the current direction could be ignored (NXP 2019). Given that the voltage of the machine tool is stable and around 230 volts, the measured power value can be calculated as:

$$P = 230 \times I_{RMS} \tag{37}$$

Table 32. Specification of the YHDC current sensor

Input current	Output voltage	Turn ratio	Work temperature
0-100A	0-50mV	100A:0.05A	-25℃ ~ +70℃

5.2.3 Piezoelectric sensor

With the purpose of the cutting force measurement, two piezoelectric sensors (specification in Table 33) have been integrated into the proposed platform, as the piezoelectric effect reflects the change in force through the change of the output charge.

Table 33. Specification of the piezoelectric sensor

Resonant frequency	Insulation resistance	Maximum input voltage	Operating temperature range
6.5±0.7KHz	100MΩ	30Vp-p	-20℃ to +70℃

5.3 Design of experiment for the acquisition platform

To assess the capability of the developed platform, an experiment has been designed to monitor the cutting tool status of the milling process by the accelerometer, piezoelectric sensor and current sensor. The acquired data are evaluated according to the Taylor's equation (Karandikar et al. 2013), which is a typical empirical formula for tool life prediction. The general form of the Taylor's equation is expressed as:

$$V_c T^n \times a_p V_f = C \tag{38}$$

where V_c denotes the cutting speed, T is the tool life in Minutes, n denotes an exponent that depends on the specific tool level and used materials, determines the slope of the tool life curve.

a_p denotes the depth of cut, V_f is the feed rate, and C is a constant that depends on the machine and workpiece material.

According to the Taylor's equation, the cutting speed (V_c), feed rate (V_f) and depth of cutting (a_p), which are highly related to the tool life, have been considered in this experiment. Based on the recommend values of milling spindle speed and feed in (Sandvik 2017), the experimental parameters employed in this work are listed in Table 34.

Table 34. Cutting parameters

No.	Spindle speed N (RPM)	Feed f (mm/tooth)
1	2000	0.0127
2	2500	0.0203
3	3500	0.0254
4	4500	0.0508

Moreover, for the milling process, the cutting speed V_c in m/min can be expressed as Equation 39, and the feed rate V_f in mm/min can be calculated as Equation 40.

$$V_c = \frac{\pi DN}{1000} \tag{39}$$

$$V_f = NZf_z \tag{40}$$

where, D denotes the tool diameter (mm, the value is 12 in this experiment), N denotes the spindle speed (RPM), Z denotes the number of cutter flutes (Z=4 in this experiment), f_z denotes the feed (mm/tooth).

In line with the 4 different parameters of spindle speed and feed, 16 machining parameter combinations have been designed through the Taguchi method, shown in Table 35. In order to achieve a good surface quality, the depth of cut has been carefully designed based on the classical method provided in (Uddeholm 2007).

Table 35. Experimental processing parameters

No.	Spindle speed N (RPM)	Cutting speed V_c (m/min)	Feed rate V_f (mm/min)	Depth of cut a_p(mm)	Width of cut (mm)
1	2000	75	102	1.5	12
2	2000	75	203	2	12
3	2000	75	356	2.5	12
4	2000	75	914	3	12
5	2500	94	102	2	12
6	2500	94	203	1.5	12
7	2500	94	356	3	12
8	2500	94	914	2.5	12
9	3500	132	102	2.5	12
10	3500	132	203	3	12
11	3500	132	356	1.5	12
12	3500	132	914	2	12
13	4500	170	102	3	12
14	4500	170	203	2.5	12
15	4500	170	356	2	12
16	4500	170	914	1.5	12

Figure 53 shows the deployment of the proposed platform. All the experiments have been carried out on an SYIL X4 CNC machine by using 4 flutes 12mm HSS milling cutter to machine aluminium blocks.

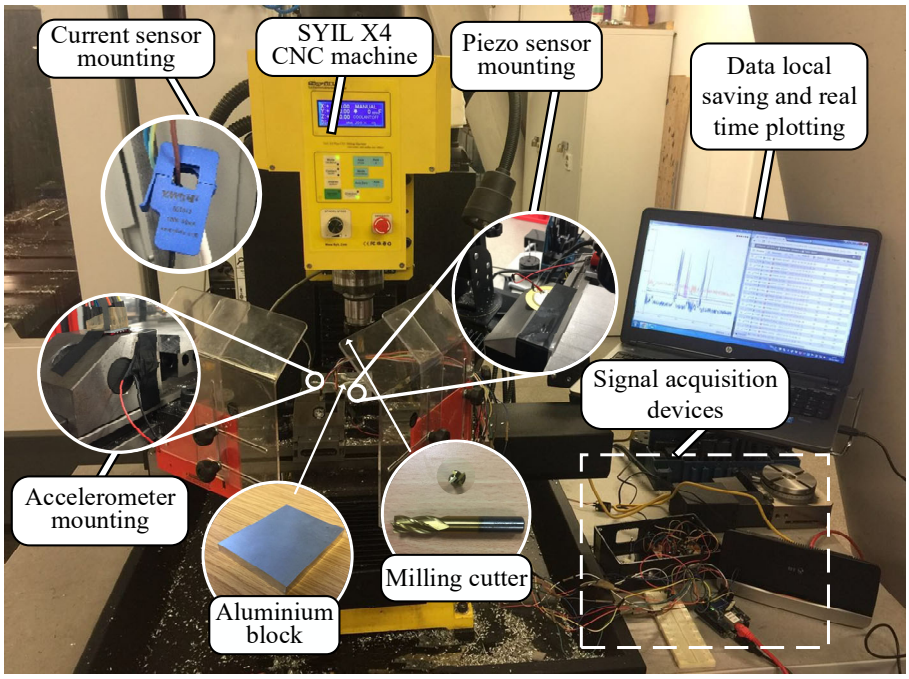


Figure 53. The system deployment of machining process monitoring

5.4 Result analysis of the acquisition platform

As shown in Figure 54, 16 slots have been machined with different parameters on 3 aluminium blocks, respectively. Throughout the experiment process, 62023 signal samples of 8 different sensors signal are acquired from 2 piezoelectric sensors, a 3-axis accelerometer, and 3 current sensors. All data have been collected by the monitoring platform steadily and stored in the developed data database.

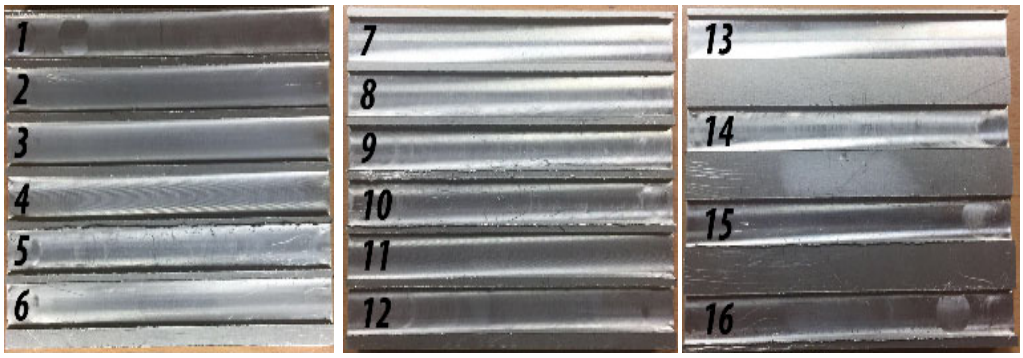


Figure 54. The workpiece after machining

5.4.1 Signal analysis based on process monitoring

The overall power and 3-axis acceleration signal captured during the experiment process are plotted in Figure 55 (a) and Figure 56 (a), respectively. From the power graph, it is noted that the highest value of power occurs at the initial stage (highlighted in the red box) of the machining, which is close to 3500W. The lowest power appears at the standby stage that is about 230W. From the time of 21:03:36 onwards, a stable and continuous processing phase starts, and the average power is 690W. For vibration signal in Figure 56(a), the acceleration values of X- and Y-axis in the initial stage are reached 0.1g and the Z-axis peak reaches nearly 6g. And the values in X- and Y-axis is about 0.05g in the continuous processing stage, and 0.4g in the Z-axis. Moreover, the values at the standby stage are 0g for X- and Y-axis and 1g for Z-axis, according to the calibration principle of the accelerometer described above, it can be judged that the accelerometer is stationary at this stage. As the signs of acceleration signal only represent the two opposite directions of acceleration, not the value, it can be observed that the trends of the power and acceleration graphs show substantial similarity. Thus, the system can be confirmed as able to collect signals to reflect the processing state.

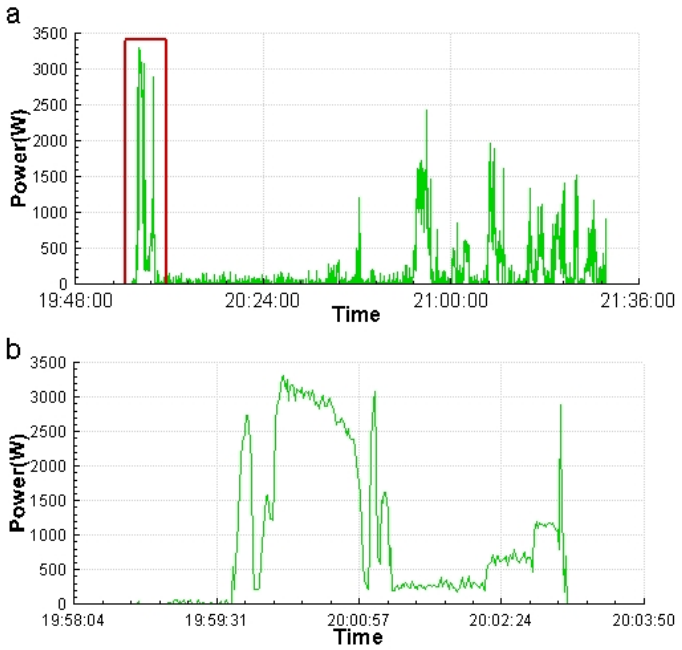


Figure 55. (a) The power signal against time; (b) The enlarged failure zone (red box in (a))

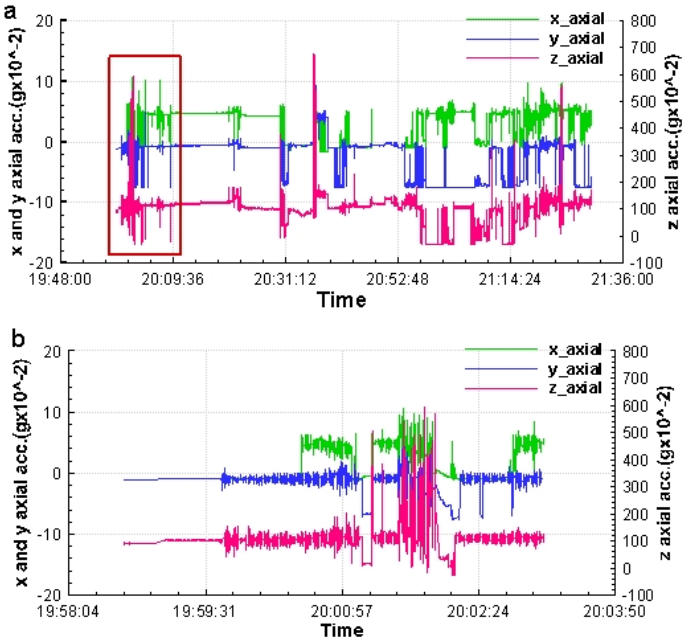


Figure 56. (a) The acceleration signal against time; (b) The enlarged failure zone (red in (a))

In practical, the abnormal peak values of the power and acceleration captured during the initial stage are caused by a failure cutting, shown in Figure 57 (a), meanwhile, it also caused the obvious wear of the cutting tool (Figure 57 (b)). The collected signal in the fault processing phase is shown in Figure 55 (b) and Figure 56 (b). Compared with normal cutting, the failure cutting shows large fluctuations in power level, and the average power of the failure cutting is about 1288W, and the

energy consumption is 0.42kwh, which is 3.5 times bigger than the normal machining with the same parameter. In addition, the similar results can be seen from Figure 56 (b), the acceleration values in 3-axis change rapidly with larger amplitudes than the normal cutting, especially in Z-axis, which displays good correlation with the power signal plotting. This phenomenon demonstrates that the vibration level of the cutter in the abnormal condition far stronger than that of the normal cutting, and it is in line with the statement by Ibrahim et al. (2017) that the increase of vibration signal amplitude implies the energy increment that generated by the tool flank wear. The potential of the proposed platform for the tool wear monitoring can be proved based on these observations.

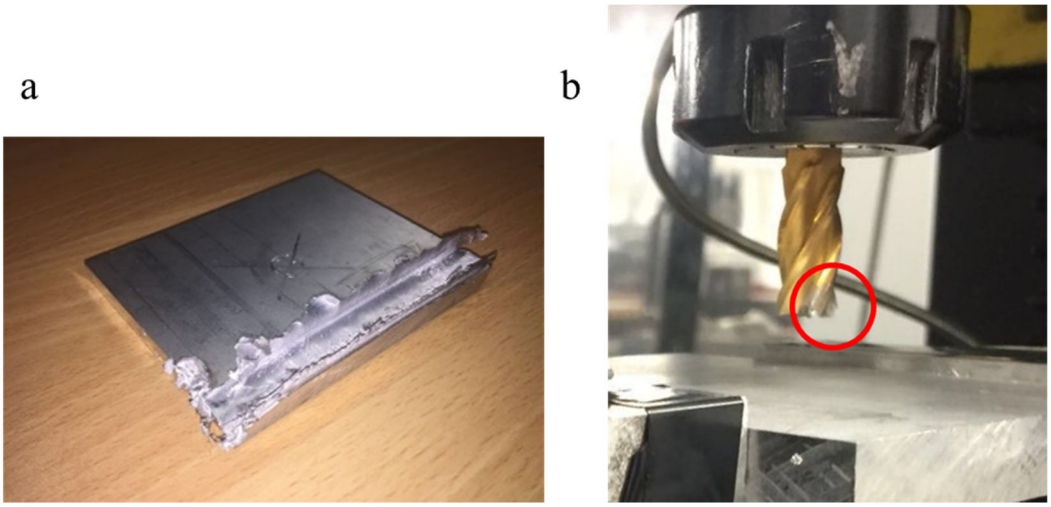


Figure 57. (a). Failed machining workpiece; (b). Worn cutting tool

In the experiment, two piezoelectric sensors are mounted on both sides of the longitudinal workpiece holder, to acquire the relevant signals reflecting the cutting force. The results are depicted in Figure 58. By comparing with the power and acceleration signals, the fluctuation of the cutting force signal is not significant, and it under a low degree of correspondence with the actual machining situation. It may because the principle of the piezoelectric sensor is more suitable for the sudden increase force, not sensitive to continuous and small changes, and the mounting location may not efficient.

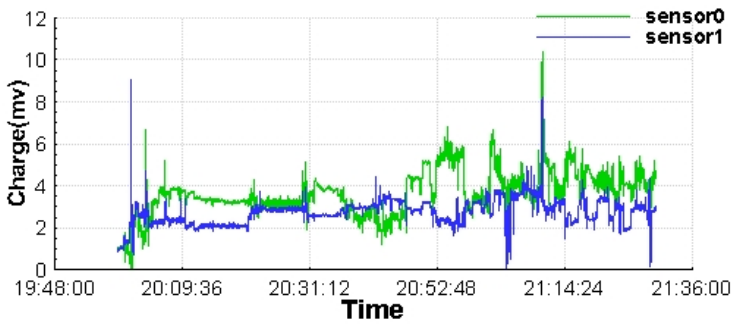


Figure 58. The signal of the piezoelectric sensor against time

5.4.2 Signal analysis based on tool life

According to the obtained data under normal machining condition, the 3-axis mean acceleration value and the power value of every slot have been averaged. The result is shown in Figure 59. It can be seen that, among the machining processes of the 16 slots, the maximum power appears at the machining of the 6th slot, it is about 1100W. The minimum value is close to 300W, which of the machining of the 3rd and 8th slot. For acceleration, the highest value reaches to 1.3g at the 4th machining and the minimum value is 0.3g at the 8th machining. Thus, the polyline trend can confirm again that the signals of the power and acceleration obtained by the system fit each other to a certain degree.

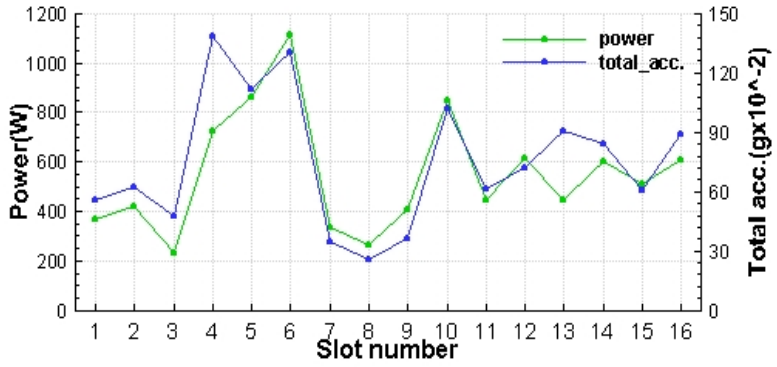


Figure 59. The average value of power and acceleration for 16 slots

In order to evaluate the correlation between the acquired signal and cutting tool life, after averaging the total acceleration and power values for each slot, the average values and parameter factors have been adopted to conduct the Taguchi analysis, since the Taguchi orthogonal method has been employed to design the parameter combination within cutting speed, feed rate and depth of cutting. The main effects plot for means is obtained as shown in Figure 60. It can be found out that the result is consistent with the finding of Taylor and other studies (Mukhopadhyay et al. 2012). The cutting speed performs a significant effect both on the acceleration and power, followed by the feed rate and the depth of cutting. Thus, it presents that the data collected by the proposed system has the feasibility to establish the relationship between parameters and tool life.

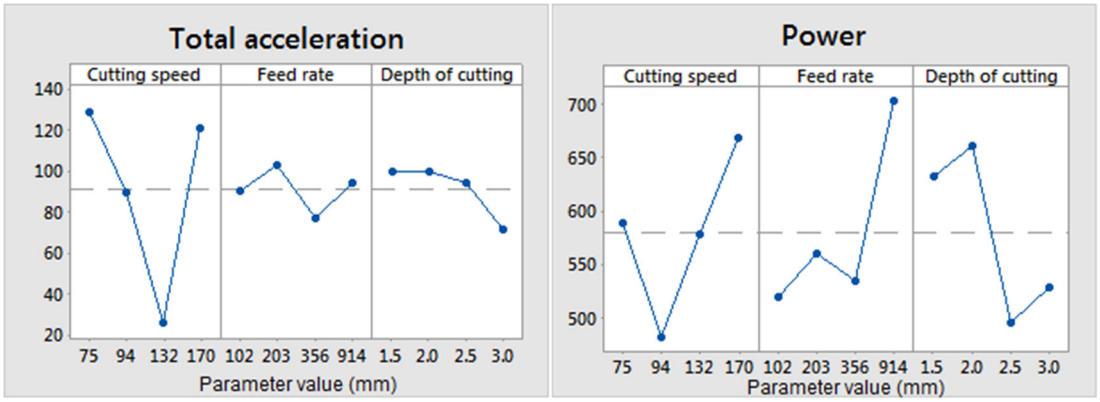


Figure 60. Main effects plot for means of the acceleration and power signal

Besides, to further assess the effectiveness of the sensor signal for revealing the tool lifetime. Figure 61 shows the power increment during the machining of the 16 slots. It can be noticed that the power increment graph is consistent with the curve of the Taylor tool wear in Figure 62. The rapidly increasing stage of power is within 5mins of machining start, which in the interval *a* (shown in Figure 61). It matches to the initial stage of the tool wear. Then the increment gets slower, which is in the interval *b* (shown in Figure 61), corresponding to the stable stage of tool wear. However, the accelerated stage of tool wear has not been reflected in the power increment graph, perhaps due to the processing time of the experiment is insufficient. Despite this, the trend of the power increment power is aligned with the Taylor curve.

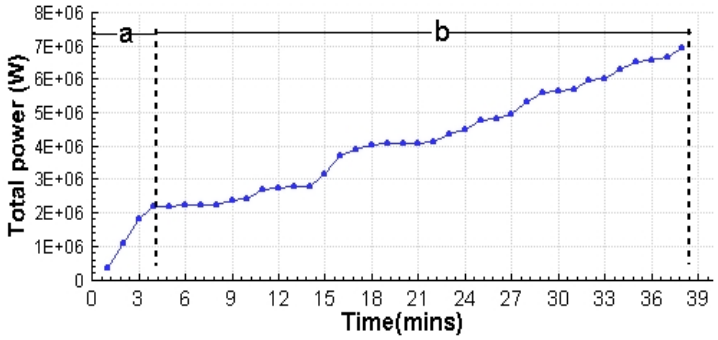


Figure 61. The superposition of the power signal

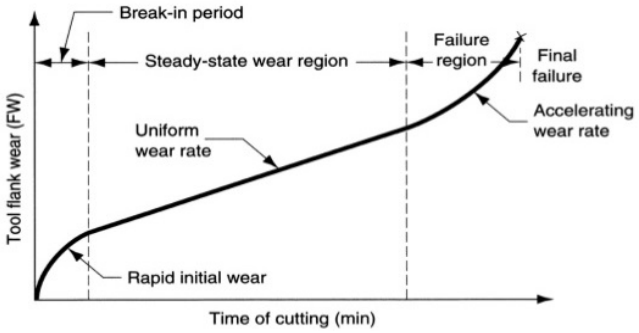


Figure 62. Taylor tool wear curve (Eslamian 2014)

5.5 Architecture of EC-based tool prognosis system

Based on the easy deployment and low-cost acquisition platform built above, the proposed EC enabled deep learning wireless cutting tool condition prognosis system is developed as shown in Figure 63. In order to realise the prompt tool wear identification and historical data-based tool RUL prediction under the circumstance of low network latency and big data processing, the system consists of three subsystems: 1) a wireless TCM system composed of multi-sensor nodes monitors the machine status in real-time during the machining processes. 2) An EC unit performs sensor signal denoising to smooth the collected sensor data, and the local data processing of converting the 1D sensor signal into 2D image format to reduce data transmission rate and enhance the data privacy. In addition, a portable deep learning model (CNN-RF model) has been embedded in this end to identify tool wear to meet real-time constrain. 3) A CC centre to manage and analyse data received from edge device through a hybrid deep learning model (hybrid CNN-LSTM model) to implement the tool RUL prediction that in low real-time demand.

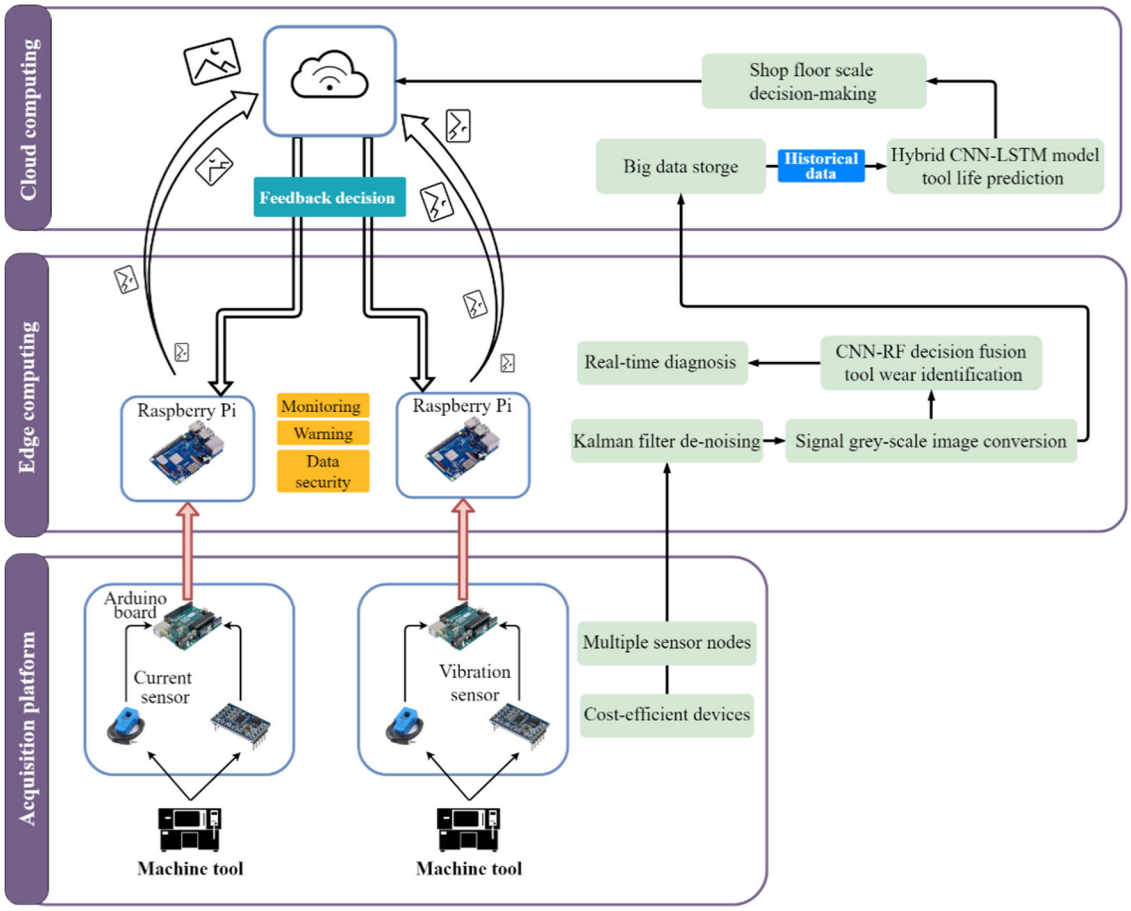


Figure 63. The architecture of the proposed system

5.5.1 Wireless TCM system

The framework of the EC-involved multi-signal acquisition platform is developed as Figure 64. A typical board Uno of Arduino (Arduino 2016) and a Wi-Fi shied (ESP8266) have been adopted for signal acquisition and flexible communication. The robust and affordable current sensor (YHDC SCT013) and an accelerometer (MMA7361) have been chosen as the signal source of power and vibration, respectively, to monitor real-time machine status.

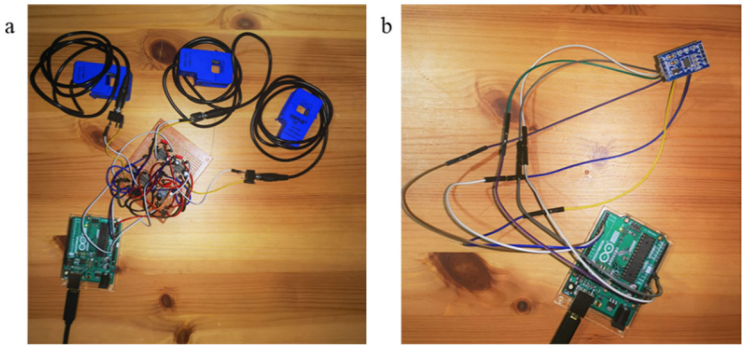


Figure 64. Wireless tool monitoring platform: (a) current node; (b) vibration node

5.5.2 Edge computing

A low computation device, Raspberry Pi (Raspberry Pi 2020), has been employed as an edge device node in this work. The framework of the EC end is shown in Figure 65. Two data processing functions, signal de-nosing and signal-to-image conversion, have been deployed on the edge computing end to enhance data processing efficiency and strength the data security and privacy. Subsequently, the converted data is processed by a lightweight decision-fusion based deep learning algorithm, which has been embedded on the edge device for the timely tool wear identification. Meanwhile, the edge computing end sends the image format data to the cloud server for storage and tool life prediction.

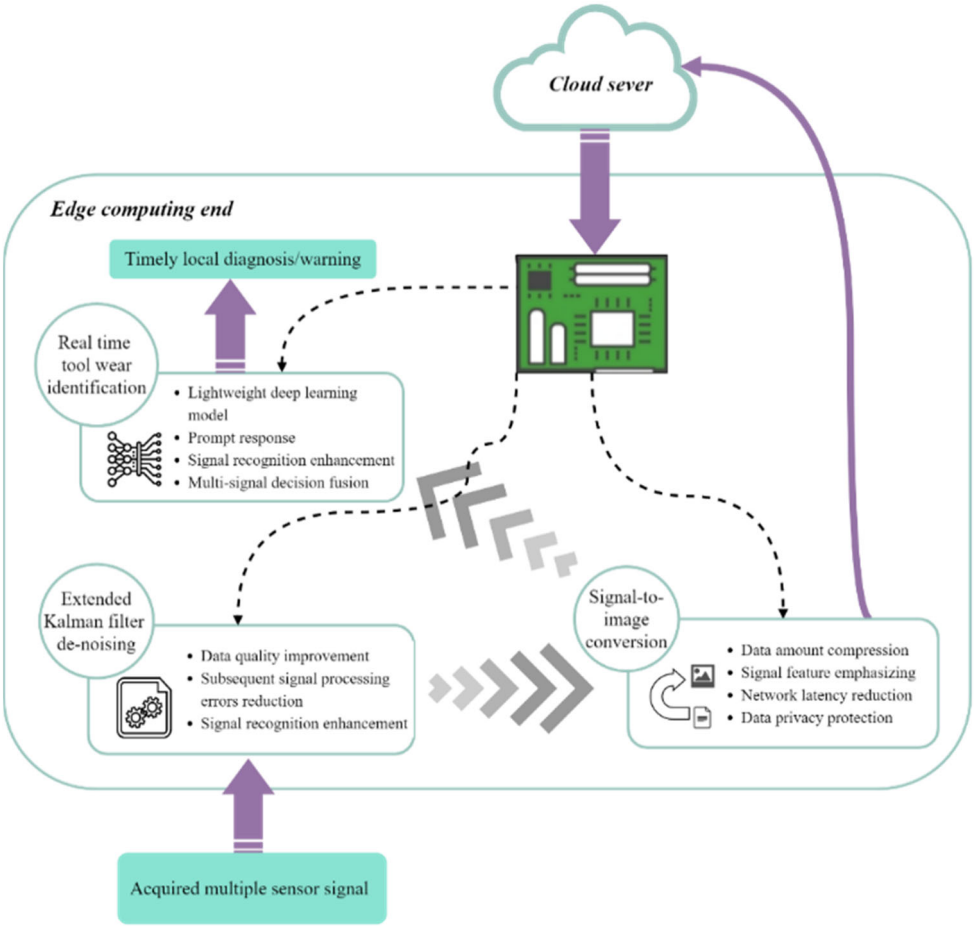


Figure 65. The framework of the edge computing end

Signal denoising

Usually, the acquired sensor data from a harsh industrial environment is full of noises. This type of sensor data directly feed to a learning model will cause signal degradation. Thus, signal denoising process is the first function that needs to be considered for the data processing process. By comparing with other data denoising approaches, Kalman filter has been demonstrated to be a most potent and efficient tool for eliminating noises from signals based on the time-domain at a low computation load, because it requires low memory and the denoising can be completed on partial time-series rather than a large amount of lengthy data (Pollreisz and TaheriNejad 2019).

The standard Kalman filter is limited to processing linear signals, which is inconsistent with the nonlinear environment of signal sources in the real world. Therefore, the extended Kalman filter (EKF) was proposed to process nonlinear signals in dynamic systems (Akram et al. 2019). The EKF has served in many signal denoising studies, e.g. ECG signal denoising (Gaamouri et al. 2019), robot localization (Ullah et al. 2020) and target tracking (Hashemi and Alfi 2019) to demonstrate an excellent signal denoise performance. In this chapter, the EKF has been selected and deployed at the edge device to process signal denoising. Figure 66 depicts the flow chart of

the EKF. First, the EKF performs the estimation phase based on the previous state of sensor signals at time step *k-1,* the estimated value $\hat{x}_k^-$ and the estimate covariance matrix P_k^- thus can be calculated, then the estimated current state is linearized by the Jacobian matrix. Secondly, combined with the measured value z_k of the sensor signal and the Kalman gain K_k at current time step *K,* the estimated state will be updated, and the state of the current time step is redefined as $\hat{x}_k$ and P_k, simultaneously, the P_k is used for the estimation of next time step. Finally, after the recursive execution, the updated sensor signal over the entire length of time is the denoised signal.

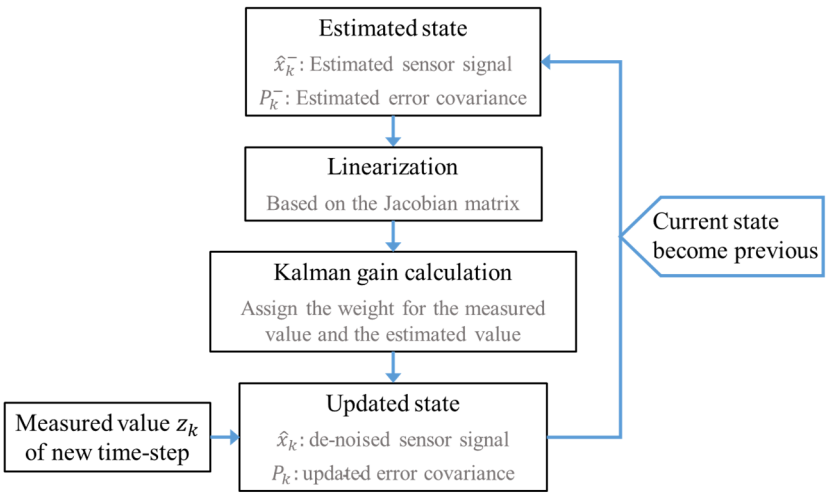


Figure 66. Flow chart of the extended Kalman filter denoising

Based on the EKF algorithm, the estimation signal samples x_k and the observation signal sample z_k collected by sensors can be calculated from Equation 41 and 42, respectively.

$$x_k = f(x_{k-1}) + \omega_k \tag{41}$$

$$z_k = h(x_k) + v_k \tag{42}$$

where ω_k denotes the process noise (e.g., from the manufacturing external environment) at time k, v_k denotes the measurement noise, x_{k-1} is previous sensor signal value, $f(\cdot)$ denotes the nonlinear function of the transition matrix, that to calculate the estimation value according to the previous estimation, $h(\cdot)$ denotes the nonlinear function of measurement matrix, which calculates the estimation measurement value according to the estimation value.

In general, the process noise ω_k and measurement noise v_k are unrelated. As the measurement noise is determined by the sensor specifications, which is hard to enhance in this work. The parameters of the process noise will be adjusted to have a better filter performance. According to Equation 41 and 42, the better denoising performance can be obtained by optimising the estimated value $\hat{x}_k$ of the real signal x_k according to the measured sensor signal value $\{z_1, z_2, \cdots z_k\}$. EKF

is a two-step recursive estimation, the first stage is the estimation, which estimates the current state based on the previous state estimation:

$$\hat{x}_k^- = f(\hat{x}_{k-1}) \tag{43}$$

$$P_k^- = F \cdot P_{k-1} F^T + \omega_k \tag{44}$$

where, $\hat{x}_k^-$ is the priori estimate value at time k, $\hat{x}_{k-1}$ is the posteriori estimate value at time k-1, P_k^- is the priori estimate covariance matrix at time k (measures the estimation accuracy), F denotes the state transition matrix, P_{k-1} is the posteriori covariance matrix at time k. It is necessary to be given is, $P_k^- = E[e_k^- e_k^{-T}]$, $P_{k-1} = E[e_{k-1} e_{k-1}^T]$, $E[\cdot]$ denotes the average value, and e_k^- and e_{k-1} are the priori estimate error and posteriori estimate error, respectively, which $e_k^- = x_k - \hat{x}_k^-$, $e_{k-1} = x_{k-1} - \hat{x}_{k-1}$.

For a nonlinear sensor signal in this chapter, EKF calculates F by successively using the Jacobian matrix and first-order Taylor series to complete the linearization of the estimated value (Lin et al. 2020). The obtained F depicted as:

$$F = \frac{\partial f}{\partial x}\Big| \hat{x}_{k-1} \tag{45}$$

Next, in the second update stage of EKF, the filter optimizes the observed signal values of the current state to achieve a more precise actual value estimation. The Kalman gain K_k is used to give the relative weight of the measured signal value and the current state estimate value, to achieve the optimal estimate (Gamse 2017), its mathematical expression, the updated predicted value and covariance matrix can be expressed as:

$$K_k = P_k^- (P_k^- + v_k)^{-1} = \frac{P_k^-}{(P_k^- + v_k)^{-1}} \tag{46}$$

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - \hat{x}_k^-) \tag{47}$$

$$P_k = (1 - K_k)\, P_k^- \tag{48}$$

The two stages of the EKF are carried out alternately until all the observed signal values are processed. In this process, Kalman gain will iterates to a stable value (Walia 2018), thereby, the optimal signal value estimation is accomplished. These estimated values are regarded as the signal after denoising.

Sensor signal conversion

The commonly used approaches of signal feature processing in fault prognosis are statistical feature extraction, wavelet transform, etc. (Terrazas et al. 2018), but these methods highly rely on expert knowledge. Nowadays, deep learning is emerging as a vital technology of IoT application systems. It shows better performance at the scale of big data and assists the EC frame to be more intelligent and efficient. CNN is one of the most popular deep learning models for machine state prognosis, which have been explored recently because of its great advantage in extracting nonlinear representative information of source data. Inspired by many successful cases of CNN in the field of image recognition, the data processing method of converting one-dimensional time series data into a two-dimensional form showed its benefits. The strategy of converting the sensor signal into a 2D form is a simple and generalized operation method, and since the sensor signal is periodic, the converted 2D form will not change much along the vertical direction, more information in the original data can be retained (Allahbakhshi et al. 2019). Besides, relying on the powerful capacity of CNN to capture the features of two-dimensional data, this strategy is able to achieve high prediction accuracy. Furthermore, in the process of converting the 1D signal vector into a 2D form, the sensor signal is arranged and compressed into a 2D array. Comparing with the 1D time-series of the sensor signal, the 2D conversion effectively reduces the size of the source data by expressing the two dependent features of temporal and spatial in the form of the image at the same time (Yang et al. 2019). It has good adaptability for edge devices with limited memory, to meet more efficient data transmission in the IoT. On the other hand, while IoT-based CC lays the foundation for the application of big data, it also opens up opportunities for industrial information leakage, hence, it is necessary for data owners to assure that the information passed to the cloud server in a confidential form (Tariq et al. 2019). In line with this need, the image format conversion of the sensor signal performed at the edge end is enabled to adequately hide the original sensor signal through signal reshaping, standardization and pixelate without sacrificing valuable information.

In order to simultaneously take advantage of the excellent performance of CNN in image feature mining and effectively compress signals to enhance data confidentiality without losing inherent information. The greyscale method, which can effectively provide comprehensive 2D information of the raw data and no parameter tuning required (Wen et al. 2017), has been innovatively used in this work to convert the 1D signal into the 2D format.

The conversion steps can be described as:

Step 1: To clean the original 1D sensor signal to ensure that the signal can be rooted and divided into the equal segment as the element of each row of the 2D matrix. During the

reciprocating cutting process of the cutting tool, short-term standby signals will appear due to repositioning or tool resetting. These signals have no effect on the classification of tool wear that can be eliminated from the signal. Therefore, a window of fixed size is set to delete the standby signal sample in the source signal, to retain the maximum number of rootable samples L as a new signal sequence. The execution is shown in Figure 67.

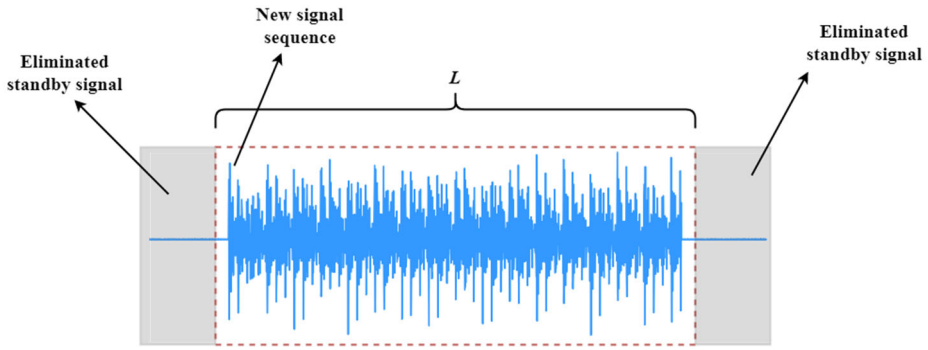


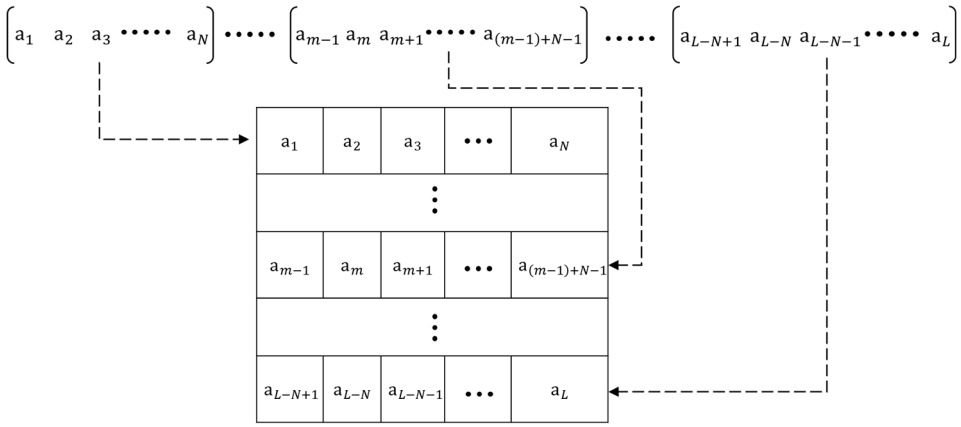
Figure 67. Sensor signal cleaning

Step 2: Based on the cleaned new sensor signal sequence, dividing the sequence into N sub-segments evenly, and each sub-segment contains N elements. This process can be depicted as below.

$$\begin{matrix}\left[a_1 \;\; a_2 \;\; a_3 \;\; \cdots\cdots \;\; a_{m-1} \;\; a_m \;\; a_{m+1} \;\; \cdots\cdots \;\; a_{L-2} \;\; a_{L-1} \;\; a_L\right] \\ \left[a_1 \;\; a_2 \;\; a_3 \;\; \cdots\cdots \;\; a_N\right] \cdots\cdots \left[a_{m-1} \;\; a_m \;\; a_{m+1} \cdots\cdots \;\; a_{(m-1)+N-1}\right] \cdots\cdots \left[a_{L-N+1} \;\; a_{L-N} \;\; a_{L-N-1} \cdots\cdots \;\; a_L\right]\end{matrix} \tag{49}$$

where, a denotes the element in the signal sequence.

Step 3: Arrange N sub-segments into a 2D matrix with the dimension of $N \times N$, the elements in the matrix corresponding to the pixel of the grey-level image. The arrangement can be represented as:



(50)

Step 4: To convert the generated matrix into greyscale pixel by normalizing the value to the pixel intensity range from 0 (black pixel) to 255 (white pixel). The normalization can be represented as:

$$P(i,j) = round\left(\frac{S(i,j) - Min(S)}{Max(S) - Min(S)} \times 255\right) \tag{51}$$

where, $i = 1, 2, \cdots N,\ j = 1, 2, \cdots N$, $S(i,j)$ denotes the sensor signal value, $P(i,j)$ denotes the pixel intensity and $round(\cdot)$ denotes the function that returns an integer. Comprehensively, the above process is demonstrated in Figure 68.

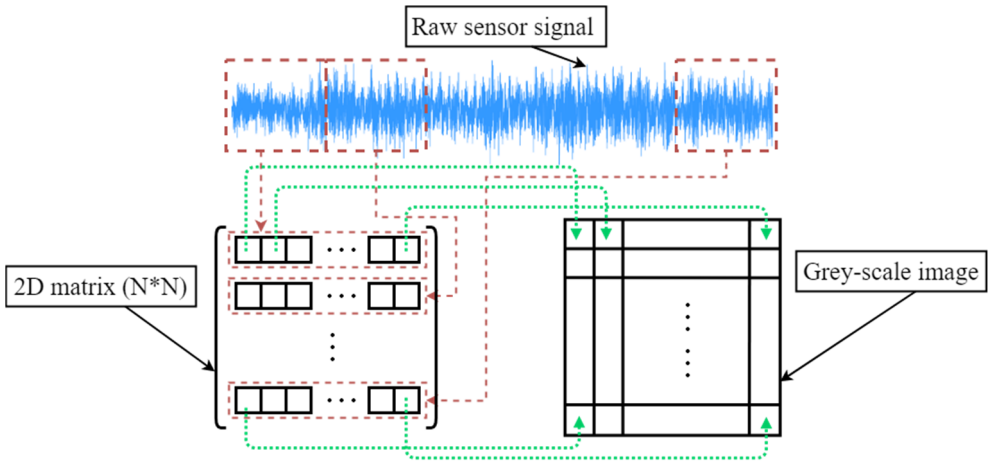


Figure 68. The conversion of a 1D signal to a 2D greyscale image

Tool wear identification

CNN is superior to other algorithms in the image local feature extraction, which is benefit from its characteristics of the local receptive field, weight sharing and sub-sampling in the spatial domain (Alshazly et al. 2019). However, the performance and expansion flexibility of 2D CNN is still limited by the size of the input data (Kim et al. 2018, Kwak et al. 2019). In addition, once

the captured feature values are not sufficient, the classification accuracy output by the softmax activation function of CNN may not be optimal (Zhou et al. 2017). To make up for the deficiencies of CNN in such situations, the integration of CNN and machine learning classifiers has been emphasised in many studies to enhance classification accuracy (e.g. Wang et al. 2019, Agarap 2019, Zhao and Liu 2019, Gallego et al. 2018). Compared to other prevalent machine learning algorithms, RF could provide excellent and consistent classification accuracy as it can maintain high execution efficiency by using all input samples as variables without requiring the intervention of parameter adjustments (Cao et al. 2013). According to (Fern´andez-Delgado et al. 2014), RF showed the best classification results among 179 various classification approaches based on using UCI dataset. On the other hand, in essence, CNN shares the weight of each neuron, so that in the process of convolution, the weight of each feature value is constant, and RF works in a similar way, which every input vector will be used equally to construct a classification model (Richmond et al. 2015). It is precisely because of these advantages that the merged application of RF and CNN has achieved gratifying results in some studies. For example, cancer mutations prediction (Agajanian et al. 2019), transportation classification (Yazdizadeh et al. 2019), semantic segmentation (Zuo and Drummond 2017), image depth estimation (Roy and Todorovic 2016) and so on. It can be seen that the integration scheme of RF and CNN will be feasible and is expected to compensate for the low accuracy of the linear classifier that comes with CNN.

In order to detect the tool wear in time for predictive maintenance, especially the failure modes such as tool deformation and abnormal wear, a lightweight hybrid CNN-RF deep learning model, which is capable of processing data on limited performance device with its simplified and effective structure, has been developed on the edge end to identify the tool wear status. Moreover, the decision-level fusion strategy has been adopted by the hybrid model, to merge the sub-decision of each sensor signal to achieve integrated decision-making, since this strategy has a high tolerance capability and will not cause interference (Roheda et al. 2018). Besides, decision-level fusion facilitates the predication model construction, which is easy to add other sensor signals into the existing system without making changes to the sub-CNN model of the remaining signals.

The proposed system is depicted in Figure 69. The workflow of the identification scheme starts from the CNN base model, which extracts features from 2D images of different sensor signals and gives individual predictions about the probability of the image being correctly classified, and then RF model gathers these generated results perform the final decision-level prediction to predict tool wear status.

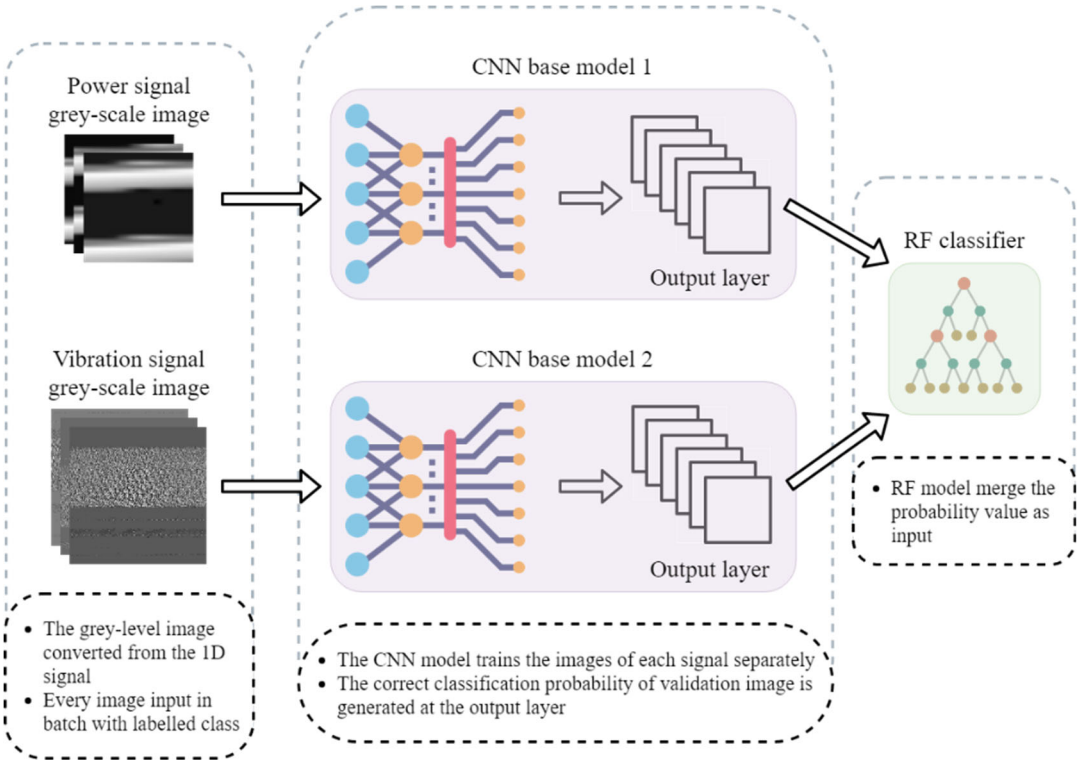


Figure 69. The architecture of decision-fusion based lightweight multi-channel CNN-RF model

Multi-channel CNN model

Nowadays, many mature and robust CNN architectures have emerged to demonstrate satisfactory performance on data analysing, for instance, LeNet-5 (Lecun et al. 1998), AlexNet (Krizhevsky et al. 2012), VGGNet (Simonyan and Zisserman 2015) and GoogLeNet (Szegedy et al. 2015). The utilisation of these architectures can effectively avoid the problems of low predictive performance caused by the deficiency of self-built models, and mostly, these successful architectures have been verified through various cases with numerous datasets. Among these successful CNN models, VGGNet and GoogLeNet both require huge computing memory and computing time, because they have complex 19 and 22 layers deep network architectures, which built for processing large-scale images (Bianco et al. 2018). LeNet-5 and AlexNet are considered as a simpler deep learning network. For LeNet-5, it includes two sets of convolutional and average pooling layers, a flatten layer, and two fully-connected layers. The results from (Elsaadouny et al. 2020) showed that it could achieve high accuracy outcome with fewer parameters. Furthermore, LeNet-5 has been proofed to have a good performance in the area of machinery status monitoring (e.g. Rahman et al. 2019, Zhang et al. 2019 and Li et al. 2020). In order to design a lightweight deep learning model that can run effectively on an edge device, the proposed CNN base model of tool wear identification is developed on the basis of LeNet-5. Figure 70 displays the architectures of the developed CNN base model for image signals.

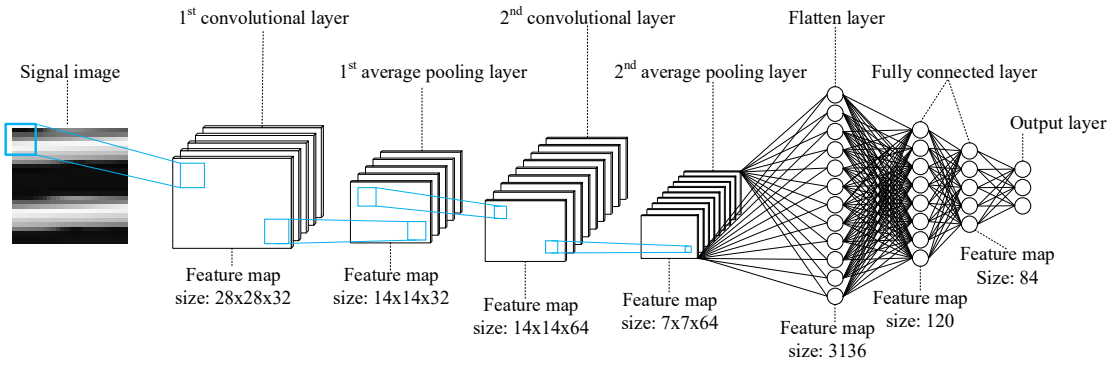


Figure 70. LeNet-5 based CNN model structure

For the CNN configuration, the image feature map $\{x_k|k = 1,2,\cdots \mathrm{N}\}$ of the sensor signal is used as input to predict the tool wear categories $\mathrm{y} = (0, 1)$. In the convolutional layer, the shared weight is used to map the input, which is the kernel, it is set as 3x3 in this layer. And the feature map output F of the convolutional layer can be obtained as:

$$F = \sum_j f(w_{i,j} * X_j + b_i) \tag{52}$$

where, $w_{i,j}$ denotes the weight for the mapping from last layer j to current layer i, X_j denotes the input from the last layer, b_i denotes the bias and $f(\cdot)$ denotes the activation functions, which is ReLU in this model.

To decrease the number of the features after the convolutional layer, the average-pooling is adopted in this model to lower the dimensional, it computes the average value in a certain size window, and then pass to the next layer (Banik et al. 2020), the kernel size of the average-pooling layer set as 2x2. On the other hand, the dropout is applied before the flatten layer to prevent the overfitting during the model training, which is a concise and effective regularization approach to weaken the sensitiveness of the CNN model to a specific neuron (Poernomo and Kang 2018).

Before outputting the classification result, the features obtained through the processing of the convolutional layer and the average-pooling layer are input to the fully connected layer to complete the mapping from feature space to the label of tool wear. The output of the fully-connected layer F_f can be described as:

$$F_f = f(w_f \cdot (F_{f-1})^T + b_f) \tag{53}$$

where, w_f denotes the weight between current and previous fully-connected layer, F_{f-1} denotes the output of the previous layer, b_f denotes the bias, $f(\cdot)$ denotes the ReLU activation functions.

Finally, the output layer of the CNN uses the sigmoid activation function to give the probability of classifying the image into a certain tool wear category, and there are two categories, namely worn and unworn. The result of the output layer can be described as:

$$P = Sigmoid(w_l \cdot (F_l)^T + b_l) \tag{54}$$

where, w_l denotes the weight between the output layer and last layer, F_l denotes the output of the previous layer, b_l denotes the bias.

RF-based decision-level fusion

RF classifier is adopted in this work to fuse the sub-prediction result of the multi-CNN models, and output comprehensive decision. It is a combination of many decision trees, and effectively prevents the deviation of prediction results caused by a single decision tree (Zhou et al. 2020). In this chapter, the classification based on RF can include as:

1. Combine the sub-decision of two CNN base models, $\mathrm{S}_a = \{X_{a1}, X_{a2}, \cdots X_{an}\}$ and $\mathrm{S}_b = \{X_{b1}, X_{b2}, \cdots X_{bn}\}$ into a new dataset $\mathrm{S} = \{\mathrm{S}_a,\ S_b\}$ as the input of RF. Where, X denotes the probability of the correct classification to the corresponding class, n denotes the number of images.
2. Based on the dataset S, use the bootstrapping of bagging method to resample S as a training dataset for the first decision tree in the RF model, and repeat this process n times to obtain a training dataset of the same size as the dataset S.
3. Randomly select m samples from the training dataset, and meet the condition $m \ll n$. Then, the m samples are spilt one by one according to the information gain to form the subsequent decision trees. Information gain G is the entropy difference before and after the splitting of samples, and the splitting stops once the gain reaches zero (Prajwala 2015). The entropy before the splitting can be represented as:

$$E(\mathrm{B}) = -\sum_{i=1}^{m} p_i \log_2 p_i \tag{55}$$

where, p_i denotes the probability of sample selection.

And the entropy after the splitting can be represented as:

$$E(\mathrm{A}) = -\sum_{i=1}^{v} \frac{D_i}{D} \cdot p_i \log_2 p_i \tag{56}$$

where, D denotes the number of samples in the training dataset, which is m. D_i denotes the number of the sample being split in the training dataset.

Therefore, the information gain can be obtained by:

$$G = E(\mathrm{B}) - E(\mathrm{A}) \tag{57}$$

4. Repeat step 2 and 3 until all decision trees in the RF model are established.
5. After the above steps, the desired RF model is trained. The aggregation of error rates of each decision tree is the prediction accuracy, which is defined as:

$$\mathrm{G} = \sum_{i=1}^{m} p_i \sum_{j \neq i}^{m} p_j = \sum_{i=1}^{m} P_i(1 - p_i) = 1 - \sum_{i=1}^{m} {p_i}^2 \tag{58}$$

where, $\sum_{i=1}^{m} p_i$ and $\sum_{j \neq i}^{m} p_j$ respectively represent the probability of correct and incorrect classification of features.

5.5.3 Cloud computing end

CC comes with powerful strength in terms of extensive-scale data processing and storage. In order to fully diagnose the tool condition to guarantee the quality of the machined component and the processing efficiency of the workshop, a tool life prediction system is developed to run on the cloud server, to prevent the workpiece surface quality reduction caused by the cutting tool gradual deterioration.

Based on the received greyscale image, the cloud end of the proposed system is responsible for the tool life prediction. In line with the same intention as the tool wear recognition of the EC end, 2D CNN is adopted as the algorithm for image processing at the cloud end. Moreover, after the images of different signals are processed in parallel by the corresponding multi-channel CNN model, the extracted feature values will be combined as the input of the subsequent LSTM based prediction model. This hybrid deep learning model greatly benefited from the superiority of CNN and LSTM algorithms, which is local feature extraction and long-term dependencies learning. Figure 71 illustrates the structure of the CNN-LSTM model.

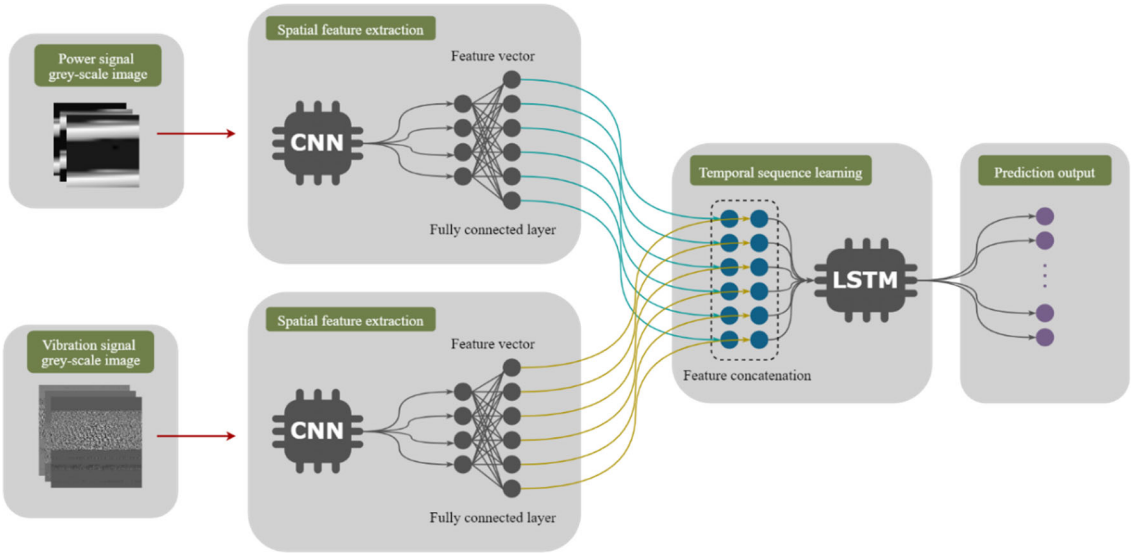


Figure 71. The architecture of hybrid multi-channel CNN-LSTM model

Multi-channel CNN model

Different from the general computing performance of EC devices, cloud server has the ability to handle increasing computing workloads. Among the four mature CNN architecture mentioned in Section 5.5.2, AlexNet is capable of achieving higher performance on the CPU with its structure of five convolutional layers, with maximum pooling layers and followed by three fully-connected layers (Krizhevsky et al. 2012), which is deeper than the LeNet5. In addition, AlexNet is considered as the inspiration of the deep learning extensive research because of its remarkable efficiency in processing the complex image (Lu et al. 2019, Shanthi and Sabeenian 2019). Moreover, AlexNet has appeared in many fields, for instance, organ classification (Igarashi et al. 2020), object detection (Bonnard et al. 2020), handwritten recognition (K.O and Poruran 2020). In this work, the CNN base model at the cloud end is developed based on the AlexNet to implement the upper-level tool life prediction, and the parameters of each layer are consistent with the CNN model at the edge end, since the input are the same. The structure of the developed AlexNet-based CNN is shown in Figure 72.

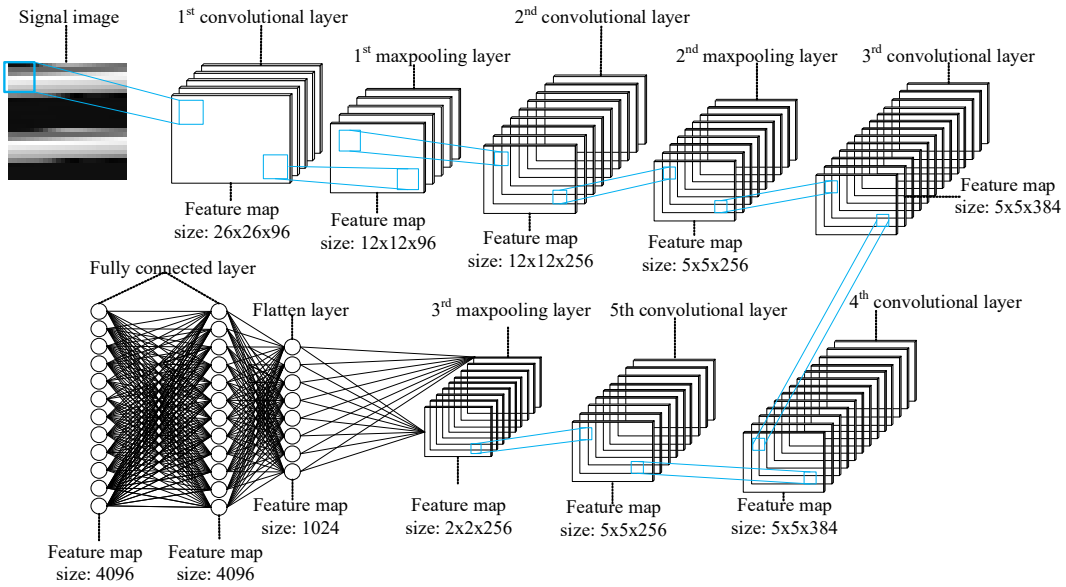


Figure 72. AlexNet based CNN model structure

LSTM prediction model

Thanks to its memory cells are competent to retain the temporal features, LSTM is superior in processing the data with time-series characteristic. The concatenated feature array $X_{m,n} = [x_{m,n}]$ of all the CNN sub-models, which x stands for the feature value, m stands for the number of features, n stands for different signals, are adopted as the input of the LSTM model. The structure of the LSTM model in the proposed system is shown in Figure 73.

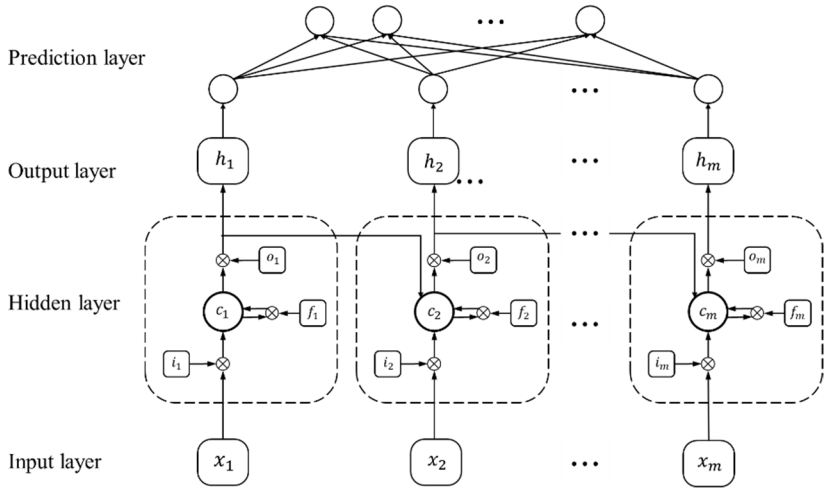


Figure 73. The structure of LSTM model

In each unit of the LSTM model, the output h_t of the current time-step t are determined by four gates, namely:

1. Input gate i_t, controls the input hidden state (output) h_{t-1} of the previous time-step $t-1$ and new input data, which is the captured feature by the CNN model. i_t can be described as:

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{59}$$

where, σ is the sigmoid activation function; W and U are variable weights and b is the bias.

2. Forget gate f_t, decides the information of i_t that to discard. It can be depicted as:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{60}$$

3. Memory gate c_t, calculates the new state of the LSTM unit based on the i_t and f_t. It can be depicted as:

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{61}$$

where, $\odot$ is the Hadamard product.

4. Output gate o_t, controls the hidden state h_t for next time-step $t+1$. It can be depicted as:

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{62}$$

During the execution of the LSTM, it recursively acquires the information from the input array $X_{m,n}$ in each time-step. The memory cell state of every LSTM unit is updated based on the information from the previous time-step c_{t-1} that filtered by the forget gate f_t, and meanwhile combine with the new fed in feature x_t, the prediction result h_t at the time-step t can be calculated under the control of output gate o_t as:

$$h_t = o_t \odot \tanh c_t \tag{63}$$

The fully-connected layer and linear regression layer are sequentially connected to the output gate of each LSTM unit. The generated features are fed into the fully-connected layer with the shape of a vector $H = \{h_1, h_2, \cdots h_l\}$, h is the output of each LSTM unit, l is the last time-step. The output of the fully-connected layer is calculated as:

$$F = \sum_{i=1} \mathrm{ReLU}(w_i \cdot h_i + b_i) \tag{64}$$

where, ReLU denotes the ReLU activation functions, w_i and b_i are the weight and bias, respectively.

Finally, the regression prediction is realised by the linear layer. And the MAE is adopted to assess the prediction performance. It can be depicted as:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\tilde{y} - y| \tag{65}$$

where, n is the training sample size; $\tilde{y}$ is the prediction value; y is the actual value.

5.6 Case study of the EC-based tool prognosis system

5.6.1 Experiment detail

The proposed EC-based TCM system was deployed on a manufacturing company in Shenzhen, China for the practical validation. The teeth machining of the ring synchronizer for 7-speed dual-clutch transmission gearbox was performed on the JCMT JCS30 CNC machine tool according to the plan. The cutting tool used in this experiment is TENX 1604 XF (Kristen+Göhrmann 2019). The ring synchronizer and insert see Figure 74.

Figure 74. The ring synchronizer and cutting tool insert

The proposed TCM system is shown in Figure 75. The MMA7361 three-axis accelerometer and three YHDC SCT013 current sensors were installed on the tool holder and the machine three-phase power supply, respectively, to monitor the vibration and power signals during the machining process.

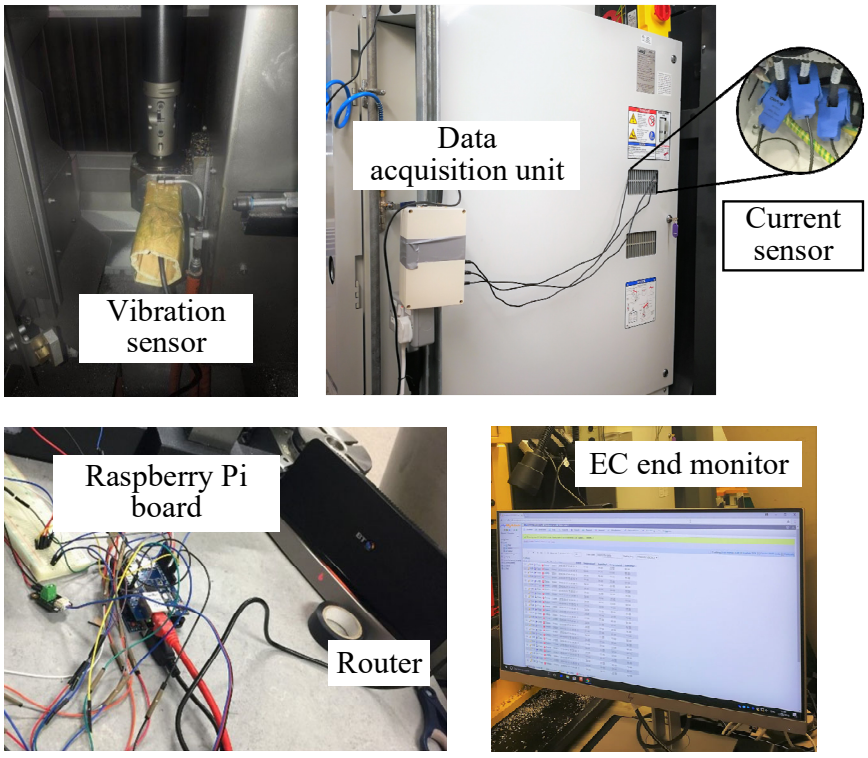


Figure 75. Sensor deployment and edge computing device

Moreover, in order to lower the randomness of the results, three groups of machining processes were carried out with the same machining parameters. The number of machined components is 37, 29 and 39 in each group, details shown in Table 36. The machining time of each group is limited to half an hour, and thus the number of machining component is different. A new cutting tool is used at the beginning of each machining group. Based on the offline quality measurement of machined components, the total machined component number and the component number machined by the worn cutting tool are recorded as Table 36.

Table 36. The condition of machined components

Machining group number	Quantity of machined component	Machined component with worn tool
1	37	26 to 37
2	29	14 to 29
3	39	11 to 39

Due to the data confidentially agreement with the company, the technical information of the components, including machining parameters was not disclosed in this chapter. During the experiment, the flank wear of three cutting tool inserts was measured offline with metrology equipment, the measurement example is shown in Figure 76. The machined components were

evaluated with the precision according to the product requirements. The tool wear status can be determined according to it.

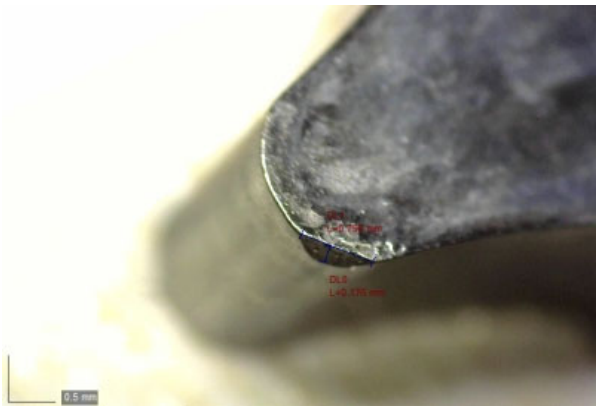


Figure 76. SEM (Scanning Electron Microscope) photograph of the tool flank wear

5.6.2 Result analysis

The acquisition frequency of vibration and power signal during the machining is set at 4000Hz and 20Hz, respectively. The partially collected power signal and the X-axis vibration signal are shown in Figure 77.

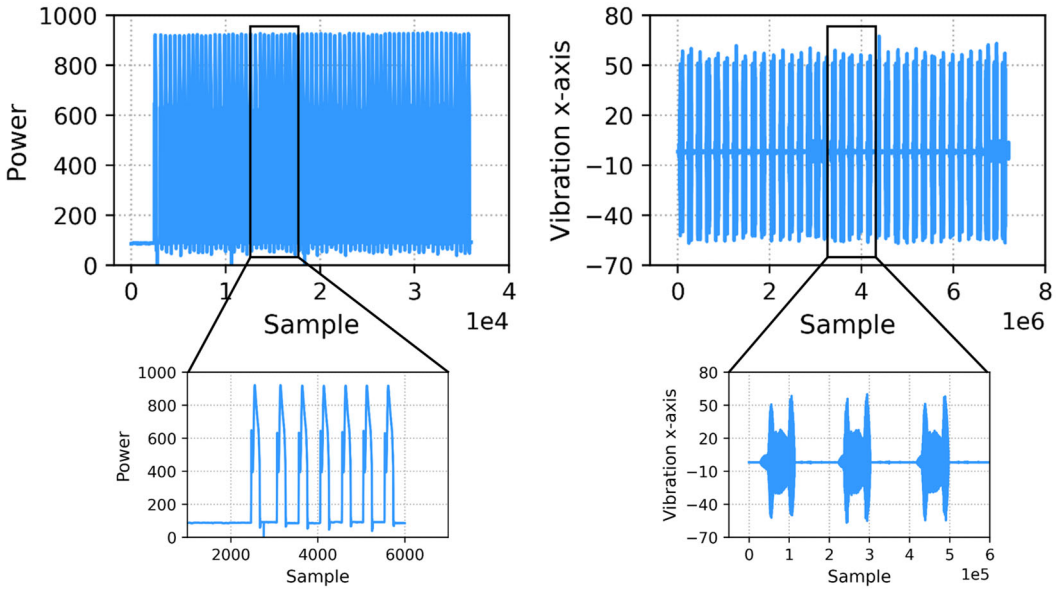


Figure 77. Acquired power and X-axis vibration signal

EKF based signal denoising

In order to reduce the deviation between the acquired signal and actual signal caused by the inevitably collected noises, the EKF is performed on the signals at the EC end to achieve the

purpose of denoising. Figure 78 shows the power and X-axis vibration signal of group 1 before and after the filtering.

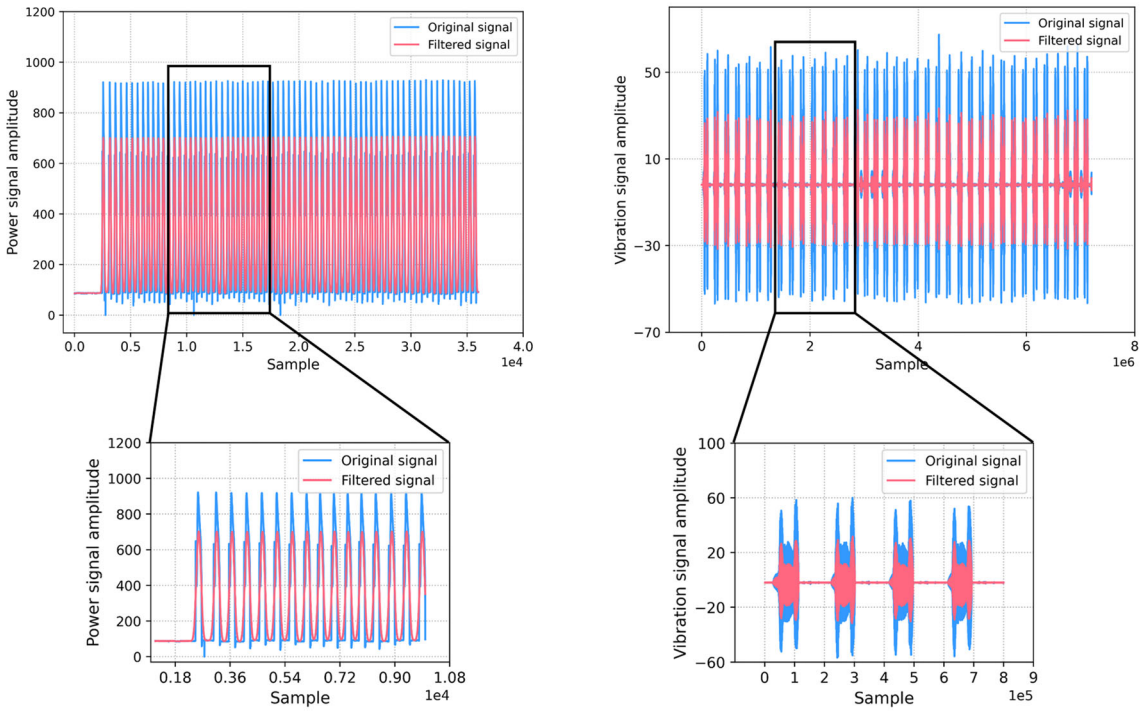


Figure 78. Original and filtered power and vibration signal

After the processing of EKF, the noise of the original power and vibration signal is effectively filtered out. Meanwhile, the denoised signal maintains the morphological characteristics of the original signal, and thus ensures the reliability of subsequent signal processing.

In order to better explain the denoised vibration signal has more significant advantages, the probability density function analysis of the original and filtered vibration signals is carried out. It reflects the probability of different signal amplitude, different vibration signals have different probability density curve shapes (Li et al. 2017), and the interference of noise information will directly lead to the changes of the probability density of the vibration signal. Furthermore, due to the influence of various unpredictable factors in the real scene during the manufacturing process, the vibration signal distribution is close to the normal distribution (Ge et al. 2018). Therefore, the normal distribution is used to characterize its probability density in this chapter, which bases on the mean value μ and standard deviation σ of the signal, and can be obtained by $P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$, X is the signal sample (Liu and Chen 2019). Figure 79 shows the probability density of power and vibration signals before and after the denoising.

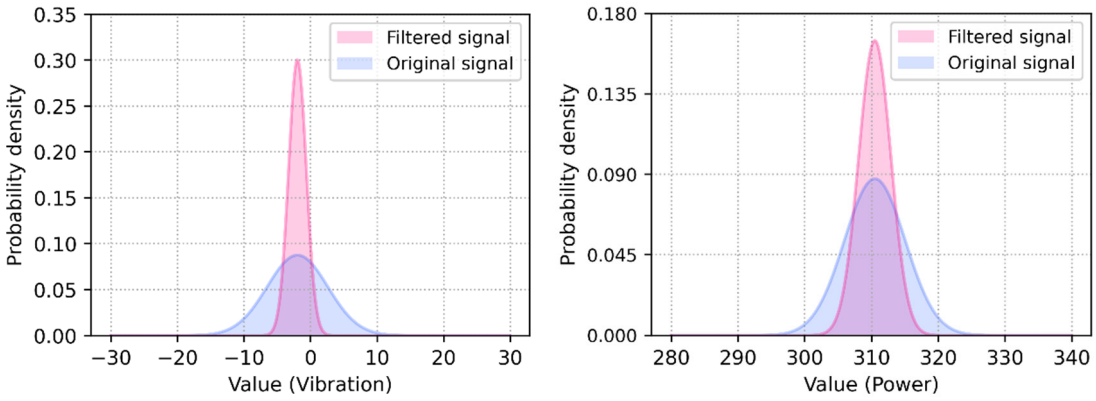


Figure 79. Probability density of the original and filtered vibration and power signal

The variable that caused the change in the shape of the probability density curve in the above figure is the standard deviation of the sensor signals, which can be used to define the 'bad' degree of the noise. The smaller standard deviation produces a higher and thinner curve shape, and indicates that the signal is stable. For the original vibration signal, the probability of the signal will not deviate from the actual value is less than 10%, while the probability of the filtered vibration signal does not deviate is raised to 30%. Moreover, along with the impact of noise on the signal increases, the curve slides down and the possibility of non-deviation is reduced, it also shows that the noise deviation of the filtered signal is less than the original signal. And the result of the power signal shows the similar trend.

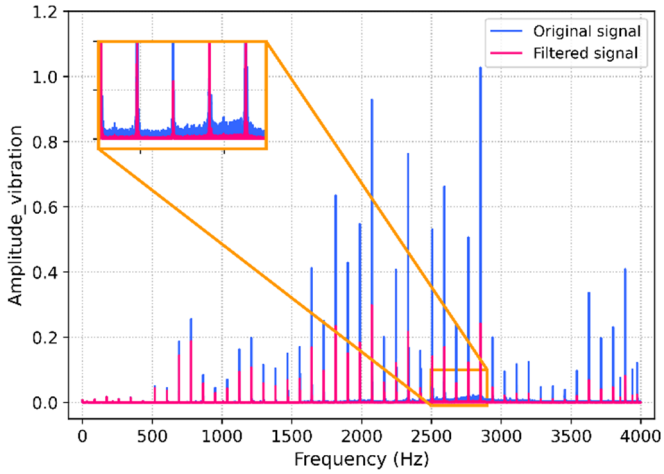


Figure 80. FFT spectrum of the original and filtered vibration signal

The frequency-domain is other standard description methods of signal denoising. FFT is used to observe the frequency component of the denoised vibration signal, as shown in Figure 80. On the whole, the frequency characteristics of the vibration signal can be imagined to be submerged by different noise frequencies, and the primary frequencies of the original signal and the denoised

signal are concentrated between 2000 Hz and 3000 Hz. Although the frequency waveform of the signal has changed after denoising, the basic shape remains, it implies that the noise is eliminated while the main characteristic information of the original signal is retained. In addition, as shown in the zoomed up area in the figure, even the noise with a small amplitude is eliminated. To sum up, the verification of the denoised vibration signal in the time and frequency-domains shows that the EKF is feasible and effective as a denoising approach in this work.

Image conversion

After the denoising processing on the EC end, the format conversion is conducted on the filtered vibration and power signal. Firstly, the signal of the two sensors is segmented in the unit of machining components. Thereby, the group 1 obtains 37 subsets of 3-axis vibration and power signals respectively, 29 subsets for group 2 and 39 subsets for group 3. The sensor signals for the 10th component of machining group 1 are displayed in Figure 81. Secondly, the size of the subset in each group is adjusted to the same size according to the signal type by eliminating the invalid value in the signal subset, for the subsequent image format conversion. There are 784 values in each power subset, and 176400 values for the vibration signal.

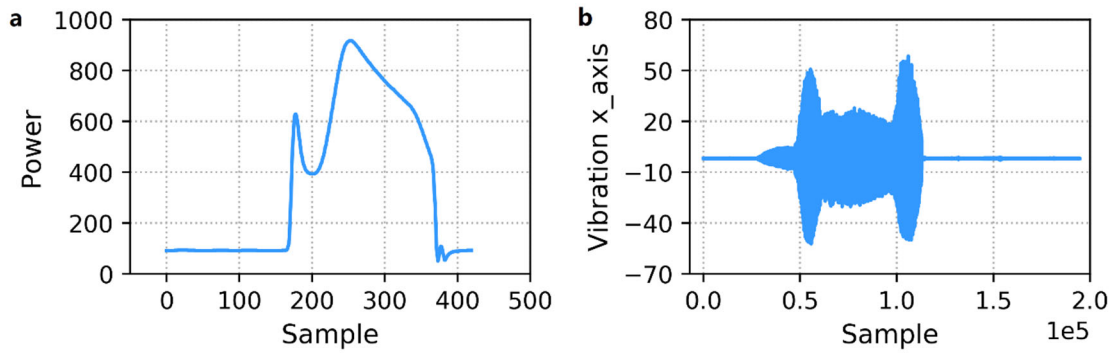


Figure 81. (a) Power signal segmentation, (b) X-axis vibration signal segmentation

According to the format conversion method illustrated in the above section, the greyscale images of two signals are constructed, which the image size of the power and vibration signal is 28x28 and 420x420, respectively. Figure 82 displays the obtained image samples of the three groups.

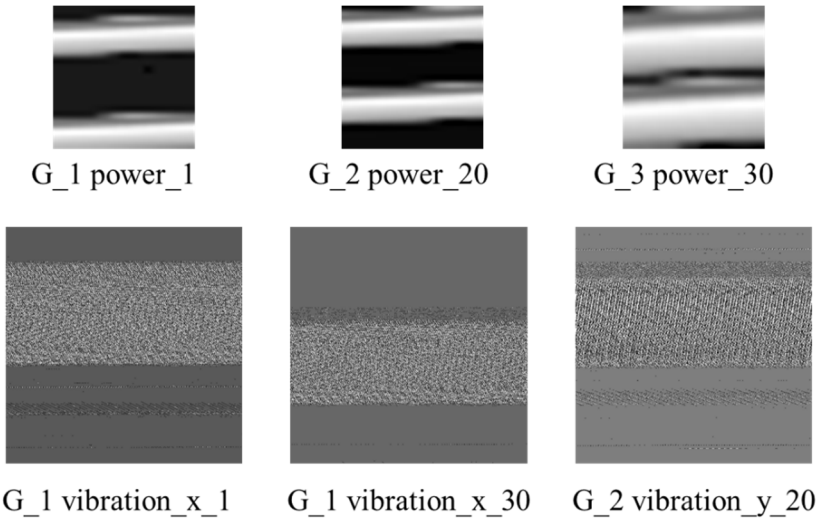


Figure 82. The images of different signals

It is not difficult to observe from the above images that the image of the power and the vibration signal present different pattern, and as the machining proceeds, the images of the same signal also show the difference, it indicates the potential for the tool wear classification. Eventually, the greyscale image of each signal subsets was generated corresponding to the number of tool cuts in each group. In three groups, there are a total of 105 power signal images and 315 vibration signal images were obtained, which also achieves the 89% data compression from 1D signal to 2D image (the average size of each power and vibration signal image is 1KB and 130KB), Figure 83 shows the comparison of the dataset size between the raw signal and signal image.

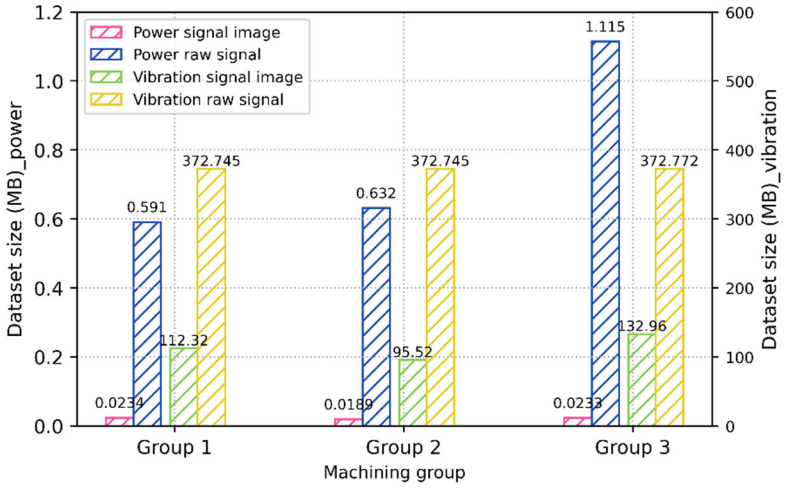


Figure 83. Signal dataset size comparison before and after the image conversion

Furthermore, as an EC device with limited computation capacity, the maximum memory of Raspberry Pi is 1G (Raspberry Pi 2020). In order to verify that this device can provide sufficient memory for data processing, the memory usage during the image conversion has been monitored.

The results show that the maximum memory usage of grey-level image batch conversion for power and vibration signal is: 256.1MiB (0.27GB) and 525.8MiB (0.55GB) (1MiB=0.001048GB, International Electrotechnical Commission 2010). In the light of the memory usage of the proposed image conversion not beyond the capacity of Raspberry Pi, this method is feasible to be applied for the EC end based on low-cost devices. For both power and vibration signal, the change of memory usage over time is shown in Figure 84.

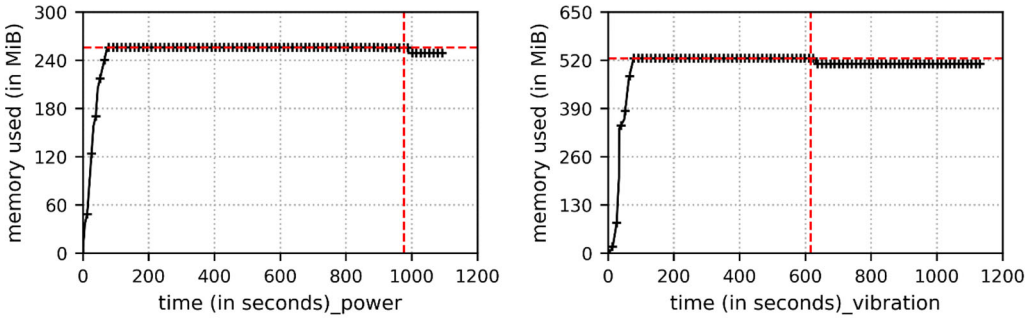
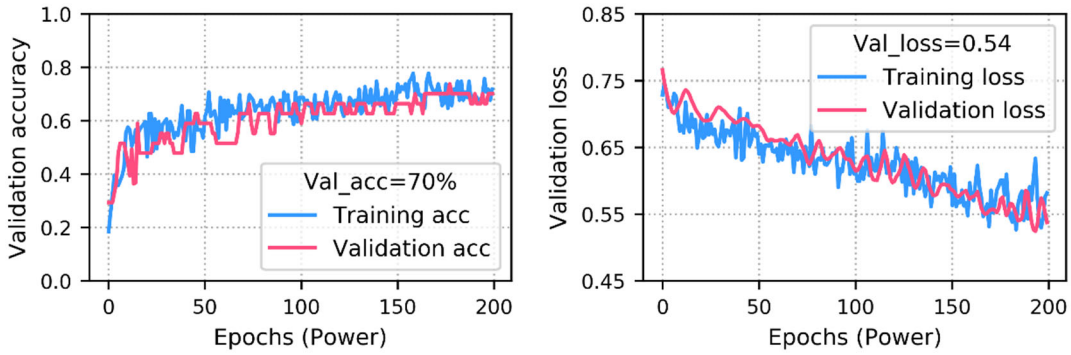


Figure 84. Memory usage of image conversion on power and vibration signal

Front-end tool wear identification

Based on the obtained signal images, the EC end executes the tool wear identification on the proposed CNN-RF model, and the performance of the model is evaluated in this section. At first, to evaluate the performance of each single CNN model, all images of three groups are input into the corresponding CNN base model according to different signals. Moreover, to ensure the rationality of evaluation, the power and vibration images of multiple same machining components are selected as the verification set, instead of a randomly generated. Figure 85 shows the performance evaluation result of the two CNN base models in multiple validation epoch cycles, which is the validation accuracy and validation loss.



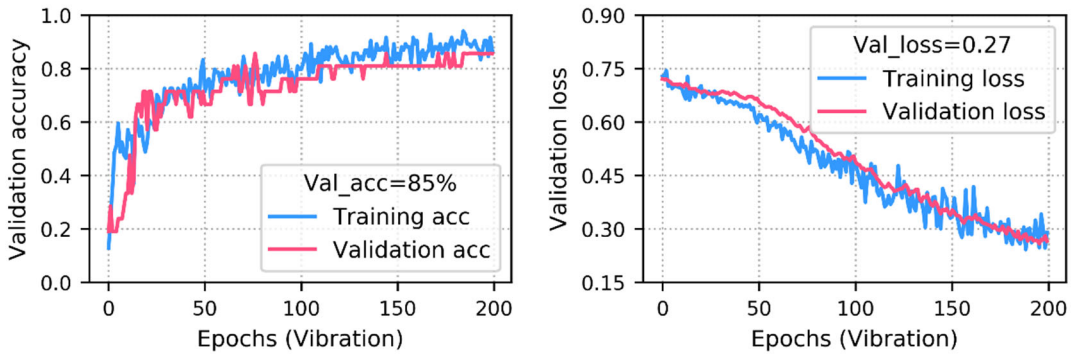


Figure 85. The evaluation result of CNN base model for the power and signal image

It can be found from the above figure, the prediction results of two CNN base models fit the training results to an acceptable degree in terms of accuracy and loss, which the classification accuracy and loss of power signal reach 70% and 0.54, and the accuracy and loss of the vibration signal achieve 85% and 0.27, respectively. It is evident that, the different signals have a different effect on the prediction performance, the accuracy of tool wear identification through vibration signal is higher than that by power signal. Nevertheless, the accuracy they provide is insufficient for timely tool status recognition at the EC end, and the fusion strategy is expected to further promote the identification performance.

Next, the CNN-RF model adopts same validation set as above to perform the identification, and the RF classifier merges the output result of two CNN base models, which is the probability of each power image and vibration image correctly classified, for the final decision-making. The machining components selected for validation and the probability of the correct classification are shown in Table 37.

Table 37. The probabilities of correct classification on power and vibration images

Group and component	Probability of correct classification		Actual wear condition
	Power image	Vibration image	
Group1_12	0.598	0.503	Unworn
Group1_17	0.502	0.609	
Group1_18	0.675	0.428	
Group1_21	0.678	0.758	
Group2_2	0.268	0.325	
Group2_25	0.698	0.719	
Group2_26	0.448	0.818	
Group2_27	0.628	0.494	
Group3_3	0.394	0.639	
Group3_7	0.687	0.723	
Group1_27	0.809	0.700	Worn
Group1_30	0.717	0.698	
Group2_9	0.299	0.679	
Group2_11	0.539	0.615	
Group2_14	0.452	0.589	
Group3_10	0.596	0.668	
Group3_15	0.447	0.557	
Group3_21	0.751	0.792	
Group3_26	0.825	0.615	
Group3_29	0.718	0.758	

Take the probability value of two signal as predictor variables, and the actual tool wear category as the respond variables (‘1’ for tool worn, ‘0’ for tool unworn), the RF model executes the tool wear identification. The performance of the classification model is assessed by four evaluation parameters, which is accuracy, precision, recall and f1-score. In particular, suppose the tool unworn is a positive result and tool worn is a negative result. The accuracy represents the ratio between the correctly predicted tool wear value and the overall predicted value, and precision represents the ratio between the correct positive predicted value and the overall positive predicted value, recall represents the ratio between the correct positive predicted value and the total predicted value, and f1-score is the weighted average of precision and recall. Table 38 shows the classification performance of the 20 machining components.

Table 38. The evaluation result of RF model

Accuracy	Precision	Recall	F1-score
90.62%	94.74%	90%	92.31%

Based on all prediction results demonstrated above, it proves that the sensor data collected by the established wireless multi-sensor acquisition platform is reliable as a raw material for tool wear identification, and the integrity of the sensor signal information is still maintained to a large extent after the processing of the introduced image format conversion method. Additionally, the decision-level fusion scheme adopted at the EC end finally achieves an improvement in the classification accuracy to 90.62%, since this method compensates the information between sensor signals.

Moreover, this chapter also monitors the memory usage of the CNN-RF model to perform tool wear classification on the validation set. The maximum usage is 334.3MiB, the proposed model consumes reasonable memory, it is promising that the well-trained CNN-RF model is unhindered to perform wear recognition at the EC end. The result is shown in Figure 86.

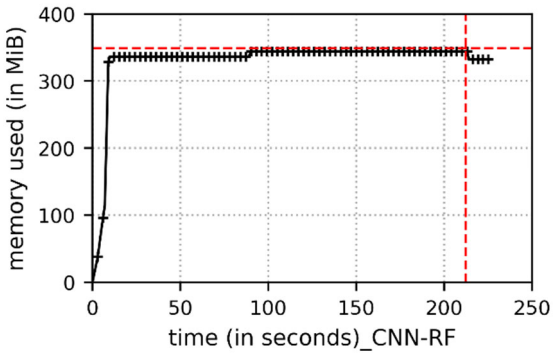


Figure 86. Memory usage of the classification validation on CNN-RF model

In order to further assess the superiority of the RF-integrated multi-channel CNN model, other four prevalent machine learning classifiers, SVM, MLP, KNN and NB classifier, are combined individually with the proposed CNN sub-model in this chapter for model performance comparison. SVM is a popular prediction algorithm. It transforms the input data into a high-dimensional space, and then obtains the maximum margin hyperplane according to the kernel function for achieving classification (Awad and Khanna 2015), it has an excellent performance in two-category classification projects (Cao et al. 2019). MLP is a simple feedforward artificial neural network. It has three-layers of one input layer, one hidden layer and one output layer. The activation function in each neuron is responsible for mapping the input data to the output layer. Although MLP has a certain degree of fault tolerance, the computation time is high (Shawky et al. 2020). KNN is an

efficient machine learning algorithm, which does not need to build a complete classification model. This method implements its classification based on the category of one or more samples nearest to the sample to be tested. If most of the adjacent samples belong to a certain category, the sample to be tested also belongs to this category (Wang et al. 2019). The NB classifier is a convenient probabilistic classifier. It assumes that the sample to be predicted is independent of each other, then the training sample set is used to create a probability model to predict the classification of new variables (Zhang and Sakhanenko 2019).

The prediction models based on these above four classifiers were developed to perform the decision-level fusion classification on the generated greyscale images, the evaluation result, in terms of accuracy, precision, recall and f1-score are shown in Table 39. Figure 87 intuitively displays each evaluation result of different classifiers.

Table 39. The evaluation results of machine learning classifiers

Decision fusion classifier	**Evaluation parameter**			
	Accuracy	Precision	Recall	F1-score
SVM	87.50%	94.44%	85.00%	89.47%
MLP	87.50%	86.36%	95.00%	90.48%
KNN	84.38%	85.71%	90.00%	87.80%
NB classifier	81.25%	93.73%	75.00%	83.33%

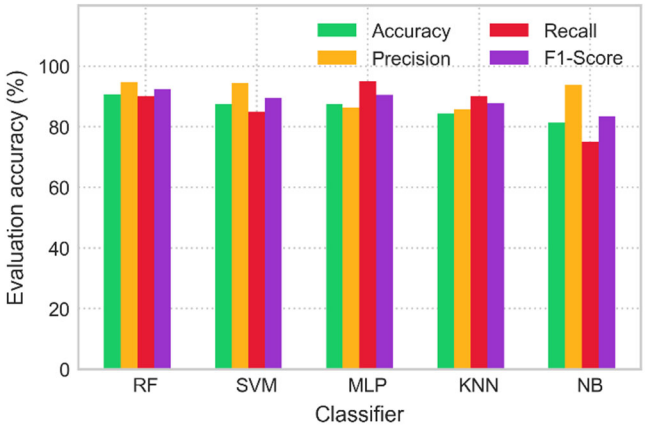


Figure 87. The evaluation accuracy of tool wear classification using different classifiers

By comparing the classification evaluation results in Table 38 and Table 39, it can be observed that the evaluation parameters of all five machine learning classifiers have reached more than 80%. Among them, RF has the best classification performance, the performance of SVM and

KNN is second only to RF, which the classification accuracy are both 87.5%. NB classifier shows the worst classification accuracy of 81.25%.

Furthermore, to assess the efficiency of each classifier, the execution time during the classification based on the validation dataset has been monitored. The result is depicted in Figure 88.

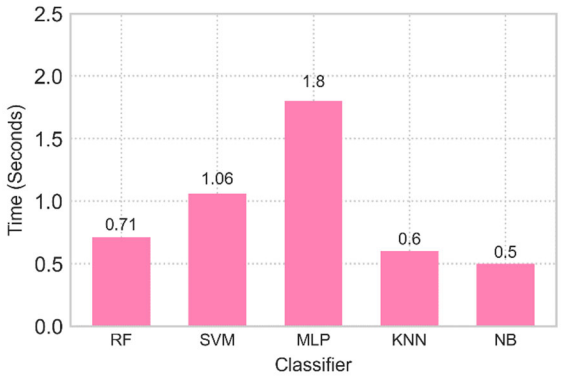


Figure 88. The computational time of different classifiers

It can be found from the above figure, among the five classifiers, MLP spends the highest time to perform classification on the validation dataset, reaching 1.8s, followed by SVM, which takes 1.06s. The least time consumption belongs to KNN and NB, which are 0.6s and 0.5s, respectively. Perhaps it is because of their relatively rapid calculations that high accuracy cannot be provided. For RF, its time consumption is moderate, which is 15.5% and 29.5% higher than KNN and NB, but 60.6% and 33% lower than MLP and SVM.

Comprehensively, by evaluating the classification accuracy and processing time of the above-mentioned classifiers, the proposed RF integrated CNN model performs better, presents a satisfactory and accurate result tool wear identification, and at the same time with the acceptable time cost.

Cloud-end prediction

In addition to the described data processing, signal format conversion and front-end tool wear identification, the EC devices also responsible for transmitting the signal image to the CC terminal. Relying on the powerful computing and storage resources of the cloud server, these data will be appropriately stored and used for the training of the cutting tool RUL prediction model, and the database is constantly updated for further optimizing the prediction model.

To verify the effectiveness of multi-channel CNN-LSTM model developed at the cloud end, the measured flank wears of each machining group are adopted as the tool life criterion in this

work. As the tool processing continues, the flank wear of three cutting tools has present the trend of sustained growth. Figure 89 shows the flank wear against the number of cuts of three cutting tools.

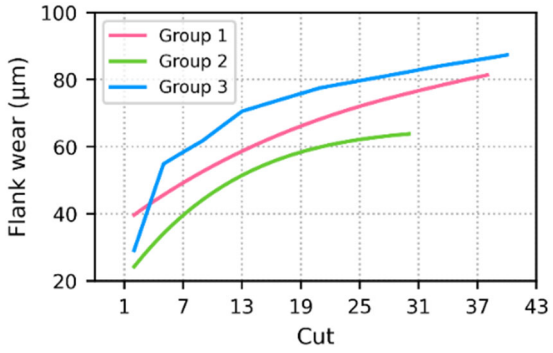
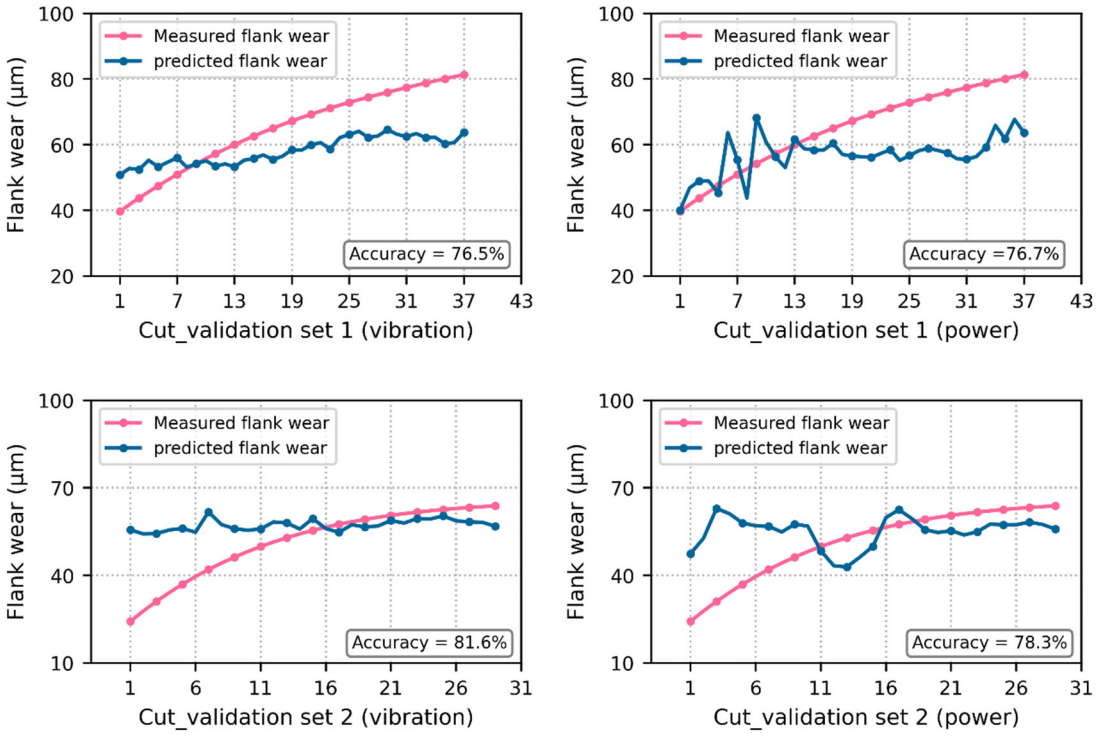


Figure 89. The flank wear of three machining tools

In view of the limited number of cuts have been performed in the case experiment in this chapter, the three machining groups are pair-wise combined as the training set and another group is the testing set, to improve the reliability of prediction result while ensuring model training performance. First, the images of the power signal and vibration signal in each generated validation set (consists of training and testing set) are separately used to conduct the flank wear prediction through the corresponding CNN base model. The prediction result is displayed in Figure 90.



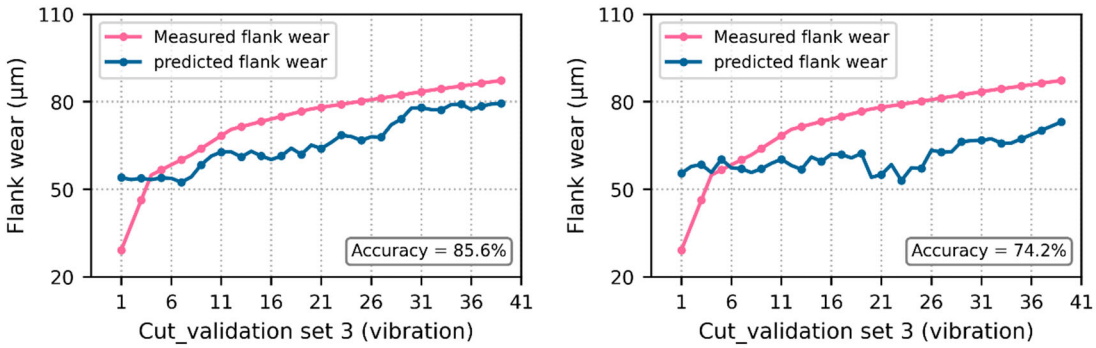


Figure 90. Predicted and measured flank wear based on different signals of each verification set

The above results demonstrate that the flank wear prediction based on the grey-level images of both sensor signal is effective, the prediction result curve is approximately fit to the actual flank wear curve. The average prediction accuracy of the power signal is 76.4%, and for vibration signal is 81.2%. The accuracy difference may because the acquisition frequency of the vibration signal is higher than the power signal.

In addition, for the purpose of further improve the prediction accuracy and improve the judgment of the RUL, the proposed CNN-LSTM model has executed the prediction via the model-based feature fusion strategy on the three validation sets. The prediction result is shown in Figure 91.

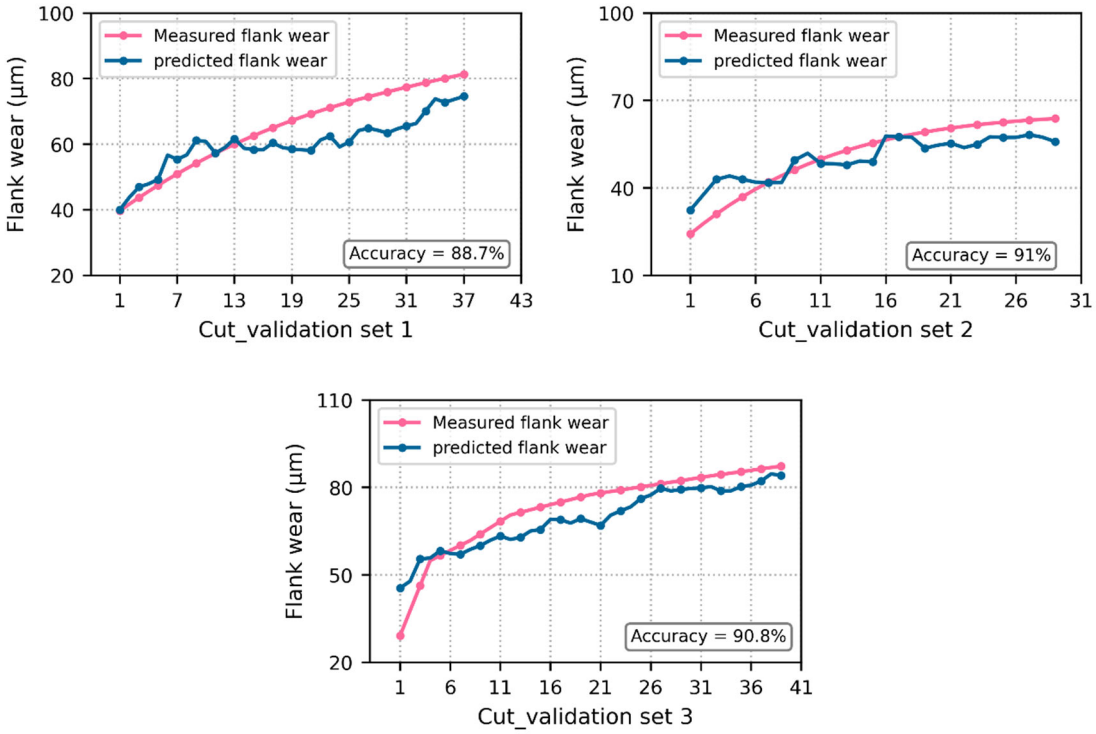


Figure 91. Predicted and measured flank wear based on CNN-LSTM model for different verification sets

The overall result displays that, benefit from the feature fusion of power and vibration signal and the long term sequence learning, the average prediction accuracy of the CNN-LSTM model reaches 90.2%, and about 15% and 10% higher than the single power signal and vibration signal. The prediction result comparison of the single CNN model and the proposed model are summarised in Table 40.

Table 40. Prediction accuracy of single CNN model and CNN-LSTM model on different validation sets

Prediction accuracy		**Validation set**					
		1		2		3	
		Training set	Testing set	Training set	Testing set	Training set	Testing set
		Group 2+3	Group 1	Group 1+3	Group 2	Group 1+2	Group 3
Single CNN model	Power signal	76.7%		78.3%		74.2%	
	Vibration signal	76.5%		81.6%		85.6%	
CNN-LSTM model		88.7%		91%		90.8%	

Finally, the cutting tool RUL can be characterised by Equation 66 based on the predicted flank wear, and the polynomial regression fitting model is employed to estimate the RUL of the cutting tools. The regression function can be described as Equation 67.

$$\mathrm{R}_p{}^i = \frac{{F_p}^i \cdot \mathrm{R}_m{}^i}{{F_m}^i} \tag{66}$$

$$R(F) = \sum_{j=0}^{k} b_j \cdot F^k \tag{67}$$

where, R_m denotes the assigned RUL based on the measured flank wear, F_p denotes the predicted flank wear, F_M denotes the measured flank wear, i denotes the number of the cuts, R denotes the RUL of a cutting tool, F denotes the flank wear, b_j denotes the regression coefficients.

Based on the three validation sets adopted in this work, the RUL estimation result is displayed in Figure 92. It demonstrates that the estimated RUL well fit the assigned RUL, and the estimation accuracy is evaluated by the coefficient of determination (R^2), which the three validation sets realise the precision of 90.6%, 92.4% and 96.7%, respectively.

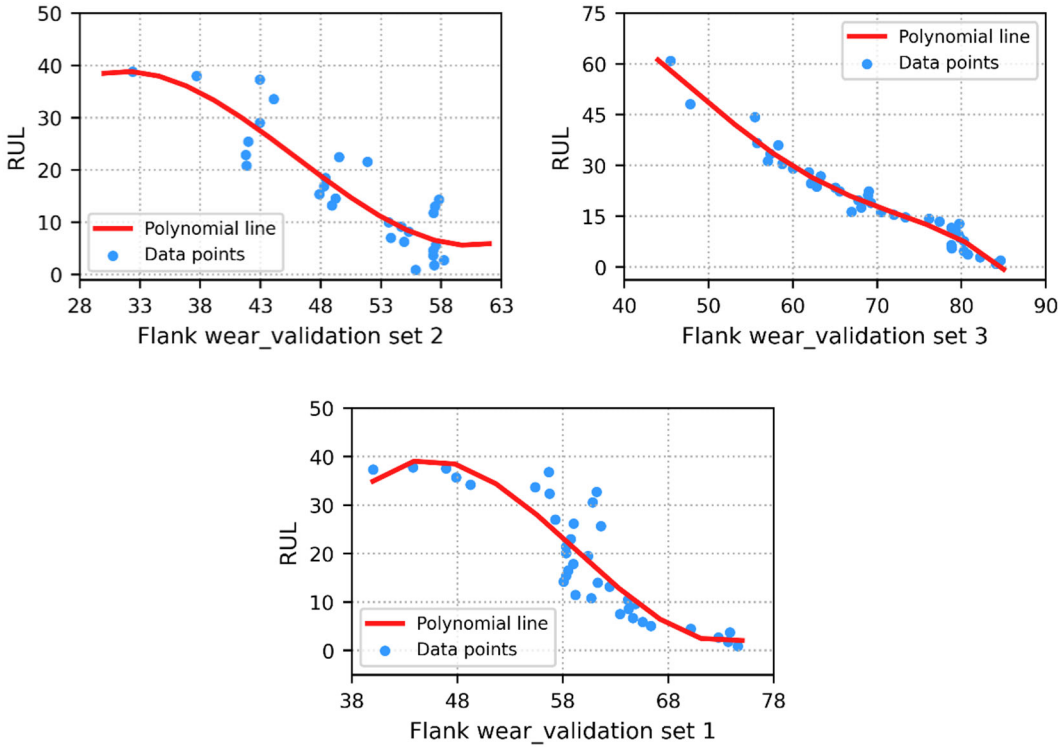


Figure 92. Predicted RUL based on the three validation sets

5.7 Summary

The rapid development of the IoT continuously promote the popularization of concept and practical implementation of cloud manufacturing, this technological innovation also brings a series of challenges that need to be alleviated, in terms of the transmission and processing of big data, efficient decision making, etc., especially for SMEs. In this chapter, a tool condition prognosis system in terms of tool wear identification and tool RUL prediction was established, which integrates the plug-and-play multi-sensor wireless acquisition platform, EC and CC. The layers of the system are based on affordable and highly open devices, the communication between each layer is achieved through Wi-Fi, and the auxiliary functions that mobile access to real-time data is provided. The workflow can be briefly described like that, the multi-sensor signal captured by the WSN is first delivered to the EC end for data processing, i.e., extended Kalman filter denoising, signal compression by 2D format conversion, and then the real-time tool wear identification is executed on the EC end through a well-trained CNN-RF model. Next, depending on the stored past and future signal images of each sensor signal from the EC end, the cloud server performs the upper-level tool RUL prediction with a hybrid multi-channel CNN-LSTM model.

The developed system was deployed in a real manufacturing environment for demonstration and verification, and satisfactory results have been achieved. The vibration and power signal of the tooth machining of the ring synchronizer has been collected stably, and then the signal

denoising and the signal format conversion have been effectively performed at the EC end, thus implement the data volume reduction of more than 89%. And, the format conversion only occupied about 50% memory capacity of the Raspberry Pi. Moreover, the tool wear identification performed by CNN-RF model at the EC end obtained the accuracy of 90.62%, and 35% of the memory capacity is consumed during the validation process. Besides, the RF is proved to perform best in comparison with SVM, MLP, KNN and NB classifier. Finally, based on the images delivered from the EC end, the designed CNN-LSTM model on the cloud server reached a global tool RUL prediction with an accuracy of 93.2%. It can be summarised that, for the intelligent EC-based IoT system, which has not been widely studied in the industrial filed, the proposed system provides a solution that integrates multiple characteristics.

Chapter 6. Conclusion and Future work

6.1 Thesis overview

In order to perform predictive maintenance of machining tools with the big data resources that comply with the development trend of the industrial IoT, thereby improving machining accuracy, and control the costs and time consumption. The primary goal of this research is to develop the efficient and SMEs-suitable TCM and prognosis system in response to Industry 4.0, which fully covers two critical assessments of tool condition, namely tool wear identification and tool RUL prediction. Regarding the methodology of this research, a series of novel methods with high efficiency and strong applicability are proposed, that correspond to heterogeneous data processing, data fusion, deep learning and EC, which are the core technologies of the intelligent IoT paradigm. Besides, this study constructed an IoT monitoring platform to realise the deployment of the low-cost multi-sensor prognosis system. In addition, the verification of the quantitative research based on real experimental data shows that the proposed methods and the established system meet the research objectives regard to tool wear identification and tool RUL prediction.

6.2 Thesis contribution

The research related to the sensor-based TCM is reviewed, and the necessary basic rationales involved in this research are introduced, which emphasises on sensor signal processing technology, application of deep learning network in the relevant monitoring field and the development of IoT monitoring system. It is found that from the literature, with the continuous upgrading of various technologies and the popularization of the intelligent industrial IoT concepts, the attention of efficient and economical monitoring systems and platforms is increasing rapidly. Both tool wear identification and tool RUL prediction have always been the interest of manufacturing research. Meanwhile, in the face of the influx of a large number of sensor signals, the data feature fusion, signal segmentation and other processing technologies are continuously improved and still demanded, to ensure more reliable prediction. In addition, deep learning that superior to the machine learning algorithms, are also emerged in the TCM field recently. By viewing plenty of existing works, the potential importance of continuing tool condition prognosis research is demonstrated, and remaining research gaps stimulate the motivation for this study.

The tool wear identification system integrating feature selection and fusion was presented. The system adopted a novel multi-level feature processing technology, in terms of signal denoising, optimal feature selection and dimensionality reduction. Firstly, the features of the collected sensor signals under the time-domain and power spectrum were extracted comprehensively, and then the

optimal size of the feature subset was obtained after recursive verification by the proposed RFECV method. Secondly, the SVM ranked the importance of each feature to the tool wear category, and finally, an optimal feature subset was determined. In addition, the Isomap fusion method, which is suitable for nonlinear data, further reduced the dimensionality of the optimal feature subset, and generated new component feature subset. By evaluating the classification performance for feature subsets from each feature processing layer on the established 1D CNN model, the proposed feature processing method is proved capable to effectively eliminate the adverse effects of redundant sensor signals on the identification accuracy, and enhance the CNN processing performance by avoiding the subjective feature selection in existing related research. Moreover, the use of 1D CNN architecture also provides a concise scheme for deep learning, thus facilities to the efficiency of the tool wear identification.

The multi-sensor data-driven tool RUL prediction system was proposed, which a Hurst exponent based signal partition methodology was developed and minimised the inaccurate tool RUL prediction caused by imbalance distribution of sensor signal. As the most common form of tool wear, flank wear during the tool life cycle was used in this work to quantify tool wear for RUL estimation. The multi-sensor signals collected during the entire tool life cycle were divided by Hurst exponent according to the fractal characteristics of different sensors, and then the signals were allocated to each stage of tool wear for segmented tool wear regression prediction. To enhance the prediction performance of tool RUL on the basis of a large amount of sensor signal, the CNN and LSTM algorithm have been ensemble as a hybrid model, and giving full play to their respective capabilities, namely the spatial feature capturing and time-series feature learning. Finally, based on an actual experimental verification, the proposed hybrid CNN-LSTM model demonstrated higher performance than other competitive deep learning models. Meanwhile, the Hurst exponent signal partitioning is also superior to other prevalent methods in the performance assessment of data processing.

A multi-nodes EC enabled IoT monitoring system with customisable configuration architectures was developed based on a verified signal acquisition platform. This system accepts customisation for specific needs, supports multi-sensor signal port, different types of sensors thus can be easily integrated into the system, which provides a foundation for industrial big data sources, and greatly reduces the developing time of new platform. In addition, due to the use of portable and cost-effective devices, the EC terminal together with the WSN can be migrated to different scenarios and be effectively deployed in the workshop to better implement distributed monitoring. The EC end embedded at the middle layer of the system, is able to optimize data resources via the signal denoising, and overcomes the inefficiency and high latency of the transmission which are common in the IoT application by converting 1D time-series signal to 2D

format, besides, benefits from the lightweight CNN-RF model configured at the EC end, the tool wear identification close to the machining frontline can be realised. Furthermore, the proposed ensemble CNN-LSTM model configured at the CC end is conducive to the big data processing based on the historical signal images, and the employed multiple sensor fusion strategies notably improve the prediction accuracy of the tool RUL.

6.3 Future work

Based on real experiments data, this study verifies the superiority and feasibility of the proposed tool wear identification method, tool RUL prediction methods, as well as the IoT prognosis system. Predictable improvements around these research may focus on:

- The state recognition scheme based on multi-layer feature processing, the RUL prediction scheme based on the signal partition, and the IoT prognosis system integrated with EC is not limited to cutting tool condition. They are low in complexity and easy to promote. Future applications in various fields are feasible, such as the prognosis of mechanical bearings and engine monitoring, etc.
- In order to further improve the effectiveness and quality of signal features to represent the integrity of the original data and continuously enhance the prediction performance, it will be worth to explore the efficient feature selection strategy with the integrated and non-stepwise structure for multiple sensors. At the same time, to perform the feature pre-processing that not only the correlation between features and tool status is considered, but also focus on features related to the site environment.
- In addition to signal fluctuations, it is hoped to find other indicators to achieve more refined and accurate signal segmentation, so as to improve the prediction efficiency by using the signals of the severe wear region of the cutting tool, without having to consume computing resources on the signal of the cutting tool in a healthy state.
- The performance of the model based on the deep learning algorithm depends on the model architecture to a certain extent. According to the different formats of the signal to be processed and the specific application scenario, the architecture optimisation of the deep learning model and the cooperation of different deep learning algorithms will be a crucial part of future work.
- For the proposed IoT monitoring system, the efficient configuration of hardware equipment and the data processing technology at the EC end will be continuously explored. In addition, as 5G technology matures, it is expected to replace Wi-Fi communications with more robust connectivity and transmission speed. At the same time,

IoT network security in the industrial environment and the encryption of the data information at the EC end will also be focused.

Bibliography

2010 PHM Society Conference Data Challenge | PHM Society (2010) available from <https://www.phmsociety.org/competition/phm/10> [13 September 2020]

Agajanian, S., Oluyemi, O., and Verkhivker, G.M. (2019) “Integration of Random Forest Classifiers and Deep Convolutional Neural Networks for Classification and Biomolecular Modeling of Cancer Driver Mutations”. Frontiers in Molecular Biosciences 6

Agarap, F.A. (2019) An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image Classification [online] available from <https://arxiv.org/pdf/1712.03541.pdf> [13 September 2020]

Aghazadeh, F., Tahan, A., and Thomas, M. (2018) “Tool Condition Monitoring Using Spectral Subtraction and Convolutional Neural Networks in Milling Process”. The International Journal of Advanced Manufacturing Technology 98 (9–12), 3217–3227

Agogino, A. and Goebel, K. (2007) BEST Lab, UC Berkeley. Milling Data Set.

Aguileta, A.A., Brena, R.F., Mayora, O., Molino-Minero-Re, E., and Trejo, L.A. (2019) “Virtual Sensors for Optimal Integration of Human Activity Data”. Sensors 19 (9), 2017

Ahmad, M.A.F., Nuawi, M.Z., Abdullah, S., Wahid, Z., Karim, Z., and Dirhamsyah, M. (2015) “Development of Tool Wear Machining Monitoring Using Novel Statistical Analysis Method, I-KazTM”. Procedia Engineering 101, 355–362

Akram, M.A., Liu, P., Tahir, M.O., Ali, W., and Wang, Y. (2019) “A State Optimization Model Based on Kalman Filtering and Robust Estimation Theory for Fusion of Multi-Source Information in Highly Non-Linear Systems”. Sensors (Basel, Switzerland) [online] 19 (7). available from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6479802/> [8 July 2020]

Alahi, Md.E.E., Mukhopadhyay, S.C., and Burkitt, L. (2018) “Imprinted Polymer Coated Impedimetric Nitrate Sensor for Real- Time Water Quality Monitoring”. Sensors and Actuators B: Chemical 259, 753–761

Ali, A., Shamsuddin, S.M., and Ralescu, A. (2015) Classification with Class Imbalance Problem: A Review [online] available from <https://www.semanticscholar.org/paper/Classification-with-class-imbalance-problem%3A-A-Ali-Shamsuddin/1e4870524f8de44d4f18c8f9f80eb797dfd25c89?p2df> [13 September 2020]

Allahbakhshi, H., Hinrichs, T., Huang, H., and Weibel, R. (2019) “The Key Factors in Physical Activity Type Detection Using Real-Life Data: A Systematic Review”. Frontiers in Physiology [online] 10. available from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6379834/> [13 September 2020]

Alshazly, H., Linse, C., Barth, E., and Martinetz, T. (2019) “Handcrafted versus CNN Features for Ear Recognition”. Symmetry 11 (12), 1493

Amin, S.U., Alsulaiman, M., Muhammad, G., Mekhtiche, M.A., and Shamim Hossain, M. (2019) “Deep Learning for EEG Motor Imagery Classification Based on Multi-Layer CNNs Feature Fusion”. Future Generation Computer Systems 101, 542–554

An, Q., Tao, Z., Xu, X., El Mansori, M., and Chen, M. (2020) “A Data-Driven Model for Milling Tool Remaining Useful Life Prediction with Convolutional and Stacked LSTM Network”. Measurement [online] 154, 107461. available from <https://www.sciencedirect.com/science/article/pii/S0263224119313284> [13 September 2020]

Arduino Uno Rev3 (2019) available from <https://store.arduino.cc/arduino-uno-rev3>

Awad, M. and Khanna, R. (2015) “Support Vector Machines for Classification”. Efficient Learning Machines 39–66

Azar, J., Makhoul, A., Barhamgi, M., and Couturier, R. (2019) “An Energy Efficient IoT Data Compression Approach for Edge Machine Learning”. Future Generation Computer Systems [online] 96, 168–175. available from <https://www.sciencedirect.com/science/article/pii/S0167739X18331716> [13 September 2020]

Balsamo, V., Caggiano, A., Jemielniak, K., Kossakowska, J., Nejman, M., and Teti, R. (2016) “Multi Sensor Signal Processing for Catastrophic Tool Failure Detection in Turning”. Procedia CIRP 41, 939–944

Bamodu, O., Osebor, F., Xia, L., Cheshmehzangi, A., and Tang, L. (2018) “Indoor Environment Monitoring Based on Humidity Conditions Using a Low-Cost Sensor Network”. Energy Procedia 145, 464–471

Banik, P.P., Saha, R., and Kim, K.D. (2020) “An Automatic Nucleus Segmentation and CNN Model Based Classification Method of White Blood Cell”. Expert Systems with Applications [online] 149, 113211. available from <https://www.sciencedirect.com/science/article/pii/S0957417420300373> [13 September 2020]

Beddar-Wiesing, S. and Bieshaar, M. (2020) “Multi-Sensor Data and Knowledge Fusion -- A Proposal for a Terminology Definition”. ArXiv:2001.04171 [Cs, Eess] [online] available from <https://arxiv.org/abs/2001.04171> [13 September 2020]

Bell, J. (2014) MACHINE LEARNING : Hands-on for Developers and Technical Professionals. S.L.: John Wiley

Benjamin, A.S., Fernandes, H.L., Tomlinson, T., Ramkumar, P., VerSteeg, C., Chowdhury, R.H., Miller, L.E., and Kording, K.P. (2018) “Modern Machine Learning as a Benchmark for Fitting Neural Responses”. Frontiers in Computational Neuroscience 12

Benkedjouh, T., Medjaher, K., Zerhouni, N., and Rechak, S. (2013) “Health Assessment and Life Prediction of Cutting Tools Based on Support Vector Regression”. Journal of Intelligent Manufacturing 26 (2), 213–223

Bhargava, K., Ivanov, S., McSweeney, D., and Donnelly, W. (2019) “Leveraging Fog Analytics for Context-Aware Sensing in Cooperative Wireless Sensor Networks”. ACM Transactions on Sensor Networks 15 (2), 1–35

Bhinge, R., Park, J., Law, K.H., Dornfeld, D.A., Helu, M., and Rachuri, S. (2017) “Towards a Generalized Energy Prediction Model for Machine Tools”. Journal of Manufacturing Science and Engineering [online] 139 (4). available from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5482378/> [13 September 2020]

Bhuiyan, M.S.H. and Choudhury, I.A. (2014) “Review of Sensor Applications in Tool Condition Monitoring in Machining”. Comprehensive Materials Processing 539–569

Bhuiyan, Md.S.H. and Choudhury, I.A. (2015) “Investigation of Tool Wear and Surface Finish by Analyzing Vibration Signals in TurningAssab-705Steel”. Machining Science and Technology 19 (2), 236–261

Bianco, S., Cadene, R., Celona, L., and Napoletano, P. (2018) “Benchmark Analysis of Representative Deep Neural Network Architectures”. IEEE Access 6, 64270–64277

Bogaerts, T., Masegosa, A.D., Angarita-Zapata, J.S., Onieva, E., and Hellinckx, P. (2020) “A Graph CNN-LSTM Neural Network for Short and Long-Term Traffic Forecasting Based on Trajectory Data”. Transportation Research Part C: Emerging Technologies [online] 112, 62–77. available from <https://www.sciencedirect.com/science/article/pii/S0968090X19309349#b0180> [13 September 2020]

Bonnard, J., Abdelouahab, K., Pelcat, M., and Berry, F. (2020) “On Building a CNN-Based Multi-View Smart Camera for Real-Time Object Detection”. Microprocessors and Microsystems [online] 77, 103177. Available from <https://www.sciencedirect.com/science/article/pii/S0141933120303446#sec0014> [13 September 2020]

Borys, P. (2020) “Long Term Hurst Memory That Does Not Die at Long Observation Times—Deterministic Map to Describe Ion Channel Activity”. Chaos, Solitons & Fractals [online] 132, 109560. available from <https://www.sciencedirect.com/science/article/pii/S096007791930517X> [13 September 2020]

Caggiano, A. (2018a) “Cloud-Based Manufacturing Process Monitoring for Smart Diagnosis Services”. International Journal of Computer Integrated Manufacturing 31 (7), 612–623

Caggiano, A. (2018b) “Tool Wear Prediction in Ti-6Al-4V Machining through Multiple Sensor Monitoring and PCA Features Pattern Recognition”. Sensors 18 (3), 823

Cao, H., Wachowicz, M., and Cha, S. (2017) Developing an Edge Computing Platform for Real-Time Descriptive Analytics [online] available from <https://ieeexplore.ieee.org/abstract/document/8258497> [13 September 2020]

Cao, J., Wang, M., Li, Y., and Zhang, Q. (2019) “Improved Support Vector Machine Classification Algorithm Based on Adaptive Feature Weight Updating in the Hadoop Cluster Environment”. PLOS ONE 14 (4), e0215136

Cao, R., Tu, W., Yang, C., Li, Q., Liu, J., Zhu, J., Zhang, Q., Li, Q.Q., and Qiu, G. (2020) “Deep Learning-Based Remote and Social Sensing Data Fusion for Urban Region Function Recognition”. ISPRS Journal of Photogrammetry and Remote Sensing [online] 163, 82–97. available from <https://www.sciencedirect.com/science/article/pii/S092427162030054X#b00653> [13 September 2020]

Cao, X.C., Chen, B.Q., Yao, B., and He, W.P. (2019) “Combining Translation-Invariant Wavelet Frames and Convolutional Neural Network for Intelligent Tool Wear State Identification”. Computers in Industry [online] 106, 71–84. available from <https://www.sciencedirect.com/science/article/pii/S0166361518304883#fig0025> [13 September 2020]

Carneiro, G., Mateus, D., Peter, L., Bradley, A., Tavares, J.M.R.S., Belagiannis, V., Papa, J.P., Nascimento, J.C., Loog, M., Lu, Z., Cardoso, J.S., and Cornebise, J. (2016) Deep Learning and Data Labeling for Medical Applications: First International Workshop, LABELS 2016, and Second International Workshop, DLMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings [online] Springer. available from <https://books.google.co.uk/books?id=p6U3DQAAQBAJ&dq=the+activation+function+increases+the+nonlinearity+between+layers> [13 September 2020]

Chang, W., Liu, Y., Xiao, Y., Yuan, X., Xu, X., Zhang, S., and Zhou, S. (2019) "A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data". Diagnostics [online] 9 (4), 178. available from <https://www.mdpi.com/2075-4418/9/4/178/htm> [13 September 2020]

Chatterjee, S., Dey, D., and Munshi, S. (2019) "Integration of Morphological Preprocessing and Fractal Based Feature Extraction with Recursive Feature Elimination for Skin Lesion Types Classification". Computer Methods and Programs in Biomedicine [online] 178, 201–218. available from <https://www.sciencedirect.com/science/article/pii/S016926071930238X> [13 September 2020]

Chaurasiya, H. (2012) "Recent Trends of Measurement and Development of Vibration Sensors". ArXiv:1209.5333 [Physics] [online] available from <https://arxiv.org/abs/1209.5333> [15 September 2020]

Chen, B., Wan, J., Celesti, A., Li, D., Abbas, H., and Zhang, Q. (2018) "Edge Computing in IoT-Based Manufacturing". IEEE Communications Magazine 56 (9), 103–109

Chen, F.C. and Jahanshahi, M.R. (2018) "NB-CNN: Deep Learning-Based Crack Detection Using Convolutional Neural Network and Naïve Bayes Data Fusion". IEEE Transactions on Industrial Electronics 65 (5), 4392–4400

Chen, J. and Ran, X. (2019) "Deep Learning With Edge Computing: A Review". Proceedings of the IEEE [online] 107 (8), 1655–1674. available from <http://www.cs.ucr.edu/~jiasi/pub/deep_edge_review.pdf> [16 December 2019]

Chen, Y., Li, H., Hou, L., and Bu, X. (2019) "Feature Extraction Using Dominant Frequency Bands and Time-Frequency Image Analysis for Chatter Detection in Milling". Precision Engineering [online] 56, 235–245. available from <https://www.sciencedirect.com/science/article/pii/S0141635918304410#bib22> [13 September 2020]

Chen, Y., Li, H., Hou, L., Wang, J., and Bu, X. (2018) “An Intelligent Chatter Detection Method Based on EEMD and Feature Selection with Multi-Channel Vibration Signals”. Measurement [online] 127, 356–365. available from <https://www.sciencedirect.com/science/article/pii/S0263224118305323#f0015> [13 September 2020]

Chen, Z., Liu, Y., and Liu, S. (2017) Mechanical State Prediction Based on LSTM Neural Netwok [online] available from <https://ieeexplore.ieee.org/abstract/document/8027963>

Chen, Z., Pang, M., Zhao, Z., Li, S., Miao, R., Zhang, Y.F., Feng, X.Y., Feng, X., Zhang, Y.X., Duan, M., Huang, L., and Zhou, F. (2019) “Feature Selection May Improve Deep Neural Networks for the Bioinformatics Problems”. Bioinformatics

Cheng, H., Chen, H., Li, Z., and Cheng, X. (2020) “Ensemble 1-D CNN Diagnosis Model for VRF System Refrigerant Charge Faults under Heating Condition”. Energy and Buildings [online] 224, 110256. available from <https://www.sciencedirect.com/science/article/pii/S0378778819338861#f0005> [13 September 2020]

Chicco, D. and Rovelli, C. (2019) “Computational Prediction of Diagnosis and Feature Selection on Mesothelioma Patient Health Records”. PLOS ONE 14 (1), e0208737

Choi, J.Y. and Lee, B. (2018) Combining LSTM Network Ensemble via Adaptive Weighting for Improved Time Series Forecasting [online] available from <https://www.hindawi.com/journals/mpe/2018/2470171/> [13 September 2020]

Corne, R., Nath, C., El Mansori, M., and Kurfess, T. (2017) “Study of Spindle Power Data with Neural Network for Predicting Real-Time Tool Wear/Breakage during Inconel Drilling”. Journal of Manufacturing Systems 43, 287–295

Costa, F., Poulichet, P., Mazaleyrat, F., and Labouré, E. (2001) “The Current Sensors in Power Electronics, a Review”. EPE Journal 11 (1), 7–18

Cuka, B. and Kim, D.W. (2017) “Fuzzy Logic Based Tool Condition Monitoring for End-Milling”. Robotics and Computer-Integrated Manufacturing 47, 22–36

Dadgari, A., Huo, D., and Swailes, D. (2018) “Investigation on Tool Wear and Tool Life Prediction in Micro-Milling of Ti-6Al-4V”. Nanotechnology and Precision Engineering [online] 1 (4), 218–225. available from <https://www.sciencedirect.com/science/article/pii/S2589554018300230> [13 September 2020]

Degenhardt, F., Seifert, S., and Szymczak, S. (2019) “Evaluation of Variable Selection Methods for Random Forests and Omics Data Sets”. Briefings in Bioinformatics [online] 20 (2), 492–503. available from <https://academic.oup.com/bib/article/20/2/492/4554516> [13 September 2020]

Dhingra, S., Madda, R.B., Patan, R., Jiao, P., Barri, K., and Alavi, A.H. (2020) “Internet of Things-Based Fog and Cloud Computing Technology for Smart Traffic Monitoring”. Internet of Things 100175

Ding, X. and He, Q. (2017) “Energy-Fluctuated Multiscale Feature Learning With Deep ConvNet for Intelligent Spindle Bearing Fault Diagnosis”. IEEE Transactions on Instrumentation and Measurement [online] 66 (8), 1926–1935. available from <https://ieeexplore.ieee.org/document/7880628> [13 September 2020]

Dong, J., Jing, B., Ma, X., Liu, H., Mo, X., and Li, H. (2018) “Hurst Exponent Analysis of Resting-State FMRI Signal Complexity across the Adult Lifespan”. Frontiers in Neuroscience [online] 12. available from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5801317/> [13 September 2020]

Duo, A., Basagoiti, R., Arrazola, P.J., Aperribay, J., and Cuesta, M. (2019) “The Capacity of Statistical Features Extracted from Multiple Signals to Predict Tool Wear in the Drilling Process”. The International Journal of Advanced Manufacturing Technology 102 (5–8), 2133–2146

Dutta, S., Pal, S.K., and Sen, R. (2016) “On-Machine Tool Prediction of Flank Wear from Machined Surface Images Using Texture Analyses and Support Vector Regression”. Precision Engineering 43, 34–42

Dymora, P. and Mazurek, M. (2019) “Anomaly Detection in IoT Communication Network Based on Spectral Analysis and Hurst Exponent”. Applied Sciences 9 (24), 5319

Dyskin, A.V., Basarir, H., Doherty, J., Elchalakani, M., Joldes, G.R., Karrech, A., Lehane, B., Miller, K., Pasternak, E., Shufrin, I., and Wittek, A. (2018) “Computational Monitoring in Real Time: Review of Methods and Applications”. Geomechanics and Geophysics for Geo-Energy and Geo-Resources 4 (3), 235–271

Ehala, J., Kaugerand, J., Pahtma, R., Astapov, S., Riid, A., Tomson, T., Preden, J.S., and Mõtus, L. (2017) “Situation Awareness via Internet of Things and In-Network Data Processing”. International Journal of Distributed Sensor Networks 13 (1), 155014771668657

Elsaadouny, M., Barowski, J., and Rolfes, I. (2020) Extracting the Features of the Shallowly Buried Objects Using LeNet Convolutional Network [online] available from <https://ieeexplore.ieee.org/document/9135701> [13 September 2020]

Elsheikh, A., Yacout, S., and Ouali, M.S. (2019) "Bidirectional Handshaking LSTM for Remaining Useful Life Prediction". Neurocomputing 323, 148–156

Eslamian, S. (2014) Handbook of Engineering Hydrology (Three-Volume Set) [online] available from <https://www.routledge.com/Handbook-of-Engineering-Hydrology-Three-Volume-Set/Eslamian/p/book/9781466552357> [13 September 2020]

Fan, K., Yin, J., Zhang, K., Li, H., and Yang, Y. (2018) "EARS-DM: Efficient Auto Correction Retrieval Scheme for Data Management in Edge Computing". Sensors (Basel, Switzerland) [online] 18 (11). available from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6263989/> [13 September 2020]

Fang, W., Zhong, B., Zhao, N., Love, P.E.D., Luo, H., Xue, J., and Xu, S. (2019) "A Deep Learning-Based Approach for Mitigating Falls from Height with Computer Vision: Convolutional Neural Network". Advanced Engineering Informatics [online] 39, 170–177. available from <https://www.sciencedirect.com/science/article/pii/S1474034618305275> [13 September 2020]

Fatehnia, M., Paran, S., Kish, S., and Tawfiq, K. (2016) "Automating Double Ring Infiltrometer with an Arduino Microcontroller". Geoderma 262, 133–139

Feng, L., Zhao, C., and Huang, B. (2019) "A Slow Independent Component Analysis Algorithm for Time Series Feature Extraction with the Concurrent Consideration of High-Order Statistic and Slowness". Journal of Process Control [online] 84, 1–12. available from <https://www.sciencedirect.com/science/article/pii/S095915241930040X#bib0125> [13 September 2020]

Ferguson, M., Bhinge, R., Park, J., Lee, Y.T., and Law, K.H. (2018) "A Data Processing Pipeline for Prediction of Milling Machine Tool Condition from Raw Sensor Data". Smart and Sustainable Manufacturing Systems 2 (1), 20180019

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., and Fernández-Delgado, A. (2014) "Do We Need Hundreds of Classifiers to Solve Real World Classification Problems?" Journal of Machine Learning Research [online] 15, 3133–3181. available from <https://jmlr.org/papers/volume15/delgado14a/delgado14a.pdf> [13 September 2020]

Fraenkel, J.R., Wallen, N.E., and Hyun, H.H. (2019) How to Design and Evaluate Research in Education. New York, Ny: Mcgraw Hill Education

Fujimura, S., Kojima, T., Okanoue, Y., Shoji, K., Inoue, M., Omori, K., and Hori, R. (2020) "Classification of Voice Disorders Using a One-Dimensional Convolutional Neural Network". Journal of Voice

Cao, G., Wang, S., Wei, B., Yin, Y., and Yang, G. (2013) "A Hybrid Cnn-Rf Method for Electron Microscopy Images Segmentation". Journal of Biomimetics Biomaterials and Tissue Engineering 18 (2)

Gaamouri, S., Salah, M.B., and Hamdi, R. (2018) "Denoising ECG Signals by Using Extended Kalman Filter to Train Multi-Layer Perceptron Neural Network". Automatic Control and Computer Sciences 52 (6), 528–538

Gallego, A.J., Pertusa, A., and Calvo-Zaragoza, J. (2018) "Improving Convolutional Neural Networks' Accuracy in Noisy Environments Using k-Nearest Neighbors". Applied Sciences [online] 8 (11), 2086. available from <https://www.mdpi.com/2076-3417/8/11/2086> [13 September 2020]

Gamse, S. (2017) "Dynamic Modelling of Displacements on an Embankment Dam Using the Kalman Filter". Journal of Spatial Science 63 (1), 3–21

Gao, Y., Gong, P., and Li, L. (2018) An End-to-End Model Based on CNN-LSTM for Industrial Fault Diagnosis and Prognosis [online] available from <https://ieeexplore.ieee.org/document/8525759> [13 September 2020]

Garbhapu, V.V. and Gopalan, S. (2017) "IoT Based Low Cost Single Sensor Node Remote Health Monitoring System". Procedia Computer Science 113, 408–415

García Plaza, E., Núñez López, P.J., and Beamud González, E.M. (2019) "Efficiency of Vibration Signal Feature Extraction for Surface Finish Monitoring in CNC Machining". Journal of Manufacturing Processes [online] 44, 145–157. available from <https://www.sciencedirect.com/science/article/pii/S1526612518311149> [13 September 2020]

García-Ordás, M.T., Alegre-Gutiérrez, E., Alaiz-Rodríguez, R., and González-Castro, V. (2018) "Tool Wear Monitoring Using an Online, Automatic and Low Cost System Based on Local Texture". Mechanical Systems and Signal Processing [online] 112, 98–112. available from <https://www.sciencedirect.com/science/article/pii/S088832701830236X> [13 September 2020]

Ge, M., Wang, J., Zhang, F., Bai, K., and Ren, X. (2018) “A Novel Fault Diagnosis Method of Rolling Bearings Based on AFEWT-KDEMI”. Entropy [online] 20 (6), 455. available from <https://www.mdpi.com/1099-4300/20/6/455> [13 September 2020]

Geramifard, O., Xu, J.X., Zhou, J.H., Li, X., and Oon, P.G. (2012) Feature Selection for Tool Wear Monitoring: A Comparative Study [online] available from <https://ieeexplore.ieee.org/document/6360911> [13 September 2020]

Ghani, J.A., Rizal, M., Nuawi, M.Z., Ghazali, M.J., and Haron, C.H.C. (2011) “Monitoring Online Cutting Tool Wear Using Low-Cost Technique and User-Friendly GUI”. Wear [online] 271 (9), 2619–2624. available from <https://www.sciencedirect.com/science/article/pii/S0043164811000871> [13 September 2020]

Ghorbani, A., Abid, A., and Zou, J. (2018) “Interpretation of Neural Networks Is Fragile”. ArXiv:1710.10547 [Cs, Stat] [online] available from <https://arxiv.org/abs/1710.10547> [13 September 2020]

Gouarir, A., Martínez-Arellano, G., Terrazas, G., Benardos, P., and Ratchev, S. (2018) “In-Process Tool Wear Prediction System Based on Machine Learning Techniques and Force Analysis”. Procedia CIRP [online] 77, 501–504. available from <https://www.sciencedirect.com/science/article/pii/S221282711831103X?via%3Dihub> [13 September 2020]

Greco, L., Ritrovato, P., and Xhafa, F. (2019) “An Edge-Stream Computing Infrastructure for Real-Time Analysis of Wearable Sensors Data”. Future Generation Computer Systems [online] 93, 515–528. available from <https://www.sciencedirect.com/science/article/pii/S0167739X18314031> [13 September 2020]

Guan, S., Pang, H., Song, W., and Kang, Z. (2018) “Cutting Tool Wear Recognition Based on MF-DFA Feature and LS-SVM Algorithm”. Www.Tcsae.Org [online] 34 (14), 61–68. available from <http://www.tcsae.org/nygcxben/ch/reader/view_abstract.aspx?file_no=20181408>

Guo, L., Gao, H., Huang, H., He, X., and Li, S. (2016) “Multifeatures Fusion and Nonlinear Dimension Reduction for Intelligent Bearing Condition Monitoring”. Shock and Vibration 2016, 1–10

Harbouche, A., Djedi, N., Erradi, M., Ben-Othman, J., and Kobbane, A. (2017) “Model Driven Flexible Design of a Wireless Body Sensor Network for Health Monitoring”. Computer Networks 129, 548–571

Hashemi, S.H. and Alfi, A. (2019) “Doppler and Bearing Tracking Using Adaptive Modified Covariance Extended Kalman Filter”. Journal of Control [online] 12 (4), 23–33. available from <http://joc.kntu.ac.ir/browse.php?a_id=493&slc_lang=en&sid=1&printcase=1&hbnr=1&hmb=1> [13 September 2020]

He, Y., Liu, Y., Shao, S., Zhao, X., Liu, G., Kong, X., and Liu, L. (2019) Application of CNN-LSTM in Gradual Changing Fault Diagnosis of Rod Pumping System [online] available from <https://www.hindawi.com/journals/mpe/2019/4203821/#conclusion> [13 September 2020]

He, Z., Shao, H., Zhong, X., and Zhao, X. (2020) “Ensemble Transfer CNNs Driven by Multi-Channel Signals for Fault Diagnosis of Rotating Machinery Cross Working Conditions”. Knowledge-Based Systems [online] 207, 106396. available from <https://www.sciencedirect.com/science/article/pii/S095070512030530X#fig4> [13 September 2020]

Herff, C. and Krusienski, D.J. (2018) “Extracting Features from Time Series”. Fundamentals of Clinical Data Science 85–100

Herzog, S., Tetzlaff, C., and Wörgötter, F. (2020) “Evolving Artificial Neural Networks with Feedback”. Neural Networks 123, 153–162

Hesser, D.F. and Markert, B. (2019) “Tool Wear Monitoring of a Retrofitted CNC Milling Machine Using Artificial Neural Networks”. Manufacturing Letters 19, 1–4

Hira, Z.M. and Gillies, D.F. (2015) A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data [online] available from <https://www.hindawi.com/journals/abi/2015/198363/>

Hochreiter, S. and Schmidhuber, J. (1997) “Long Short-Term Memory”. Neural Computation 9 (8), 1735–1780

Hosseini, M.P., Tran, T.X., Pompili, D., Elisevich, K., and Soltanian-Zadeh, H. (2020) “Multimodal Data Analysis of Epileptic EEG and Rs-FMRI via Deep Learning and Edge Computing”. Artificial Intelligence in Medicine [online] 104, 101813. available from <https://www.sciencedirect.com/science/article/pii/S0933365718306882> [13 September 2020]

Hu, J., Hong, D., Wang, Y., and Zhu, X.X. (2019) “A Comparative Review of Manifold Learning Techniques for Hyperspectral and Polarimetric SAR Image Fusion”. Remote Sensing [online] 11 (6), 681. available from <https://www.mdpi.com/2072-4292/11/6/681/htm> [13 September 2020]

Hu, Q., Qin, A., Zhang, Q., He, J., and Sun, G. (2018) “Fault Diagnosis Based on Weighted Extreme Learning Machine With Wavelet Packet Decomposition and KPCA”. IEEE Sensors Journal 18 (20), 8472–8483

Huang, C.Y., Lee, R.M., and Yang, S.K. (2016) Implement of Low Cost MEMS Accelerometers for Vibration Monitoring of Milling Process [online] available from <https://ieeexplore.ieee.org/document/7539864> [13 September 2020]

Huang, S., Cai, N., Pacheco, P.P., Narandes, S., Wang, Y., and Xu, W. (2018) “Applications of Support Vector Machine (SVM) Learning in Cancer Genomics”. Cancer Genomics & Proteomics 15 (1)

Huang, Z., Zhu, J., Lei, J., Li, X., and Tian, F. (2020) “Tool Wear Predicting Based on Multi-Domain Feature Fusion by Deep Convolutional Neural Network in Milling Operations”. Journal of Intelligent Manufacturing [online] 31 (4), 953–966. available from <https://link.springer.com/article/10.1007%2Fs10845-019-01488-7> [13 September 2020]

Ibrahim, M.R., Sreedharan, T., Fadhlul Hadi, N.A., Mustapa, M.S., Ismail, A.E., Hassan, M.F., and Tajul Arifin, A.M. (2017) “The Effect of Cutting Speed and Feed Rate on Surface Roughness and Tool Wear When Machining Machining D2 Steel”. Materials Science Forum 909, 80–85

Igarashi, S., Sasaki, Y., Mikami, T., Sakuraba, H., and Fukuda, S. (2020) “Anatomical Classification of Upper Gastrointestinal Organs under Various Image Capture Conditions Using AlexNet”. Computers in Biology and Medicine 124, 103950

International Electrotechnical Commission (2010) IEC 60050 - International Electrotechnical Vocabulary - Details for IEV Number 112-01-27: “Binary Prefix” [online] available from <http://www.electropedia.org/iev/iev.nsf/display?openform&ievref=112-01-27> [13 September 2020]

Intrieri, E., Gigli, G., Gracchi, T., Nocentini, M., Lombardi, L., Mugnai, F., Frodella, W., Bertolini, G., Carnevale, E., Favalli, M., Fornaciai, A., Alavedra, J., Mucchi, L., Nannipieri, L., Rodriguez-Lloveras, X., Pizziolo, M., Schina, R., Trippi, F., and Casagli, N. (2018) “Application of an Ultra-Wide Band Sensor-Free Wireless Network for Ground Monitoring”. Engineering Geology [online] 238, 1–14. available from <https://www.sciencedirect.com/science/article/pii/S0013795217305215> [13 September 2020]

Iqbal, R., Butt, T.A., Shafiq, M.O., Talib, M.W.A., and Umar, T. (2018) “Context-Aware Data-Driven Intelligent Framework for Fog Infrastructures in Internet of Vehicles”. IEEE Access [online] 6, 58182–58194. available from <https://ieeexplore.ieee.org/abstract/document/8488344> [13 September 2020]

Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.A. (2019) “Deep Learning for Time Series Classification: A Review”. Data Mining and Knowledge Discovery

ISO 8688-2:1989 [online] available from <https://www.iso.org/standard/16092.html> [13 September 2020]

Jacso, A., Matyasi, G., and Szalay, T. (2019) “The Fast Constant Engagement Offsetting Method for Generating Milling Tool Paths”. The International Journal of Advanced Manufacturing Technology 103 (9–12), 4293–4305

Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., Van de Walle, R., and Van Hoecke, S. (2016) “Convolutional Neural Network Based Fault Detection for Rotating Machinery”. Journal of Sound and Vibration [online] 377, 331–345. available from <https://www.sciencedirect.com/science/article/pii/S0022460X16301638> [16 April 2019]

Jia, Z., Guangchang, D., Feng, C., Xiaodan, X., Chengming, Q., and Lin, L. (2019) “A Deep Learning Fusion Recognition Method Based On SAR Image Data”. Procedia Computer Science 147, 533–541

Jing, B., Long, Z., Liu, H., Yan, H., Dong, J., Mo, X., Li, D., Liu, C., and Li, H. (2017) “Identifying Current and Remitted Major Depressive Disorder with the Hurst Exponent: A Comparative Study on Two Automated Anatomical Labeling Atlases”. Oncotarget [online] 8 (52), 90452–90464. available from <https://www.oncotarget.com/article/19860/text/> [13 September 2020]

Jing, L., Wang, T., Zhao, M., and Wang, P. (2017) “An Adaptive Multi-Sensor Data Fusion Method Based on Deep Convolutional Neural Networks for Fault Diagnosis of Planetary Gearbox”. Sensors [online] 17 (2), 414. available from <https://www.mdpi.com/1424-8220/17/2/414/htm> [13 September 2020]

K.O, M.A. and Poruran, S. (2020) “OCR-Nets: Variants of Pre-Trained CNN for Urdu Handwritten Character Recognition via Transfer Learning”. Procedia Computer Science 171, 2294–2301

Kagermann, H., Wahlster, W. and Helbig, J. (2020) Recommendations For Implementing The Strategic Initiative INDUSTRIE 4.0 [online] available from <https://www.semanticscholar.org/paper/Recommendations-for-implementing-the-strategic-4.0-Henning/80d4b3eccd6951cf6dba61f1b33889a0edbf0407> [13 September 2020]

Kaddar, B., Fizazi, H., and Mansouri, D.E.K. (2019) Convolutional Neural Network Features Selection Based On Analysis Of Variance [online] available from <https://ieeexplore.ieee.org/abstract/document/8988139> [16 September 2020]

Kahya, M.A., Altamir, S.A., and Algamal, Z.Y. (2019) “Improving Firefly Algorithm-Based Logistic Regression for Feature Selection”. Journal of Interdisciplinary Mathematics 22 (8), 1577–1581

Karam, S., Centobelli, P., D’Addona, D.M., and Teti, R. (2016) “Online Prediction of Cutting Tool Life in Turning via Cognitive Decision Making”. Procedia CIRP 41, 927–932

Karami, M., McMorrow, G.V., and Wang, L. (2018) “Continuous Monitoring of Indoor Environmental Quality Using an Arduino-Based Data Acquisition System”. Journal of Building Engineering 19, 412–419

Karandikar, J.M., Abbas, A.E., and Schmitz, T.L. (2013) “Tool Life Prediction Using Random Walk Bayesian Updating”. Machining Science and Technology 17 (3), 410–442

Khaire, U.M. and Dhanalakshmi, R. (2019) “Stability of Feature Selection Algorithm: A Review”. Journal of King Saud University - Computer and Information Sciences

Khan, A., Sohail, A., Zahoora, U., and Qureshi, A.S. (2020) “A Survey of the Recent Architectures of Deep Convolutional Neural Networks”. Artificial Intelligence Review

Khan, S., Parkinson, S., and Qin, Y. (2017) “Fog Computing Security: A Review of Current Applications and Security Solutions”. Journal of Cloud Computing 6 (1)

Khoa, V.V. and Takayama, S. (2018) “Wireless Sensor Network in Landslide Monitoring System with Remote Data Management”. Measurement [online] 118, 214–229. available from <https://www.sciencedirect.com/science/article/pii/S0263224118300022> [13 September 2020]

Khodabandehlou, H., Pekcan, G., and Fadali, M.S. (2018) “Vibration-based Structural Condition Assessment Using Convolution Neural Networks”. Structural Control and Health Monitoring e2308

Kim, T.Y. and Cho, S.B. (2019) “Predicting Residential Energy Consumption Using CNN-LSTM Neural Networks”. Energy 182, 72–81

Kim, Y., Kwak, G.H., Lee, K.D., Na, S.I., Park, C.W., and Park, N.W. (2018) “Performance Evaluation of Machine Learning and Deep Learning Algorithms in Crop Classification: Impact of Hyper-Parameters and Training Sample Size”. Korean Journal of Remote Sensing [online] 34 (5), 811–827. available from <https://www.koreascience.or.kr/article/JAKO201832853257278.page> [13 September 2020]

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., and Inman, D.J. (2019) “1D Convolutional Neural Networks and Applications: A Survey”. ArXiv:1905.03554 [Cs, Eess] [online] available from <https://arxiv.org/abs/1905.03554> [13 September 2020]

Knight, M.I. and Nunes, M.A. (2018) “Long Memory Estimation for Complex-Valued Time Series”. Statistics and Computing 29 (3), 517–536

Kong, D., Chen, Y., Li, N., Duan, C., Lu, L., and Chen, D. (2019) “Relevance Vector Machine for Tool Wear Prediction”. Mechanical Systems and Signal Processing 127, 573–594

Kong, Z., Cui, Y., Xia, Z., and Lv, H. (2019) “Convolution and Long Short-Term Memory Hybrid Deep Neural Networks for Remaining Useful Life Prognostics”. Applied Sciences 9 (19), 4156

Krishnakumar, P., Rameshkumar, K., and Ramachandran, K. (2018) Feature Level Fusion of Vibration and Acoustic Emission Signals in Tool Condition Monitoring Using Machine Learning Classifiers [online] available from <https://www.phmsociety.org/sites/phmsociety.org/files/phm_submission/2017/ijphm_18_008.pdf> [13 September 2020]

Krishnakumar, P., Rameshkumar, K., and Ramachandran, K.I. (2015) “Tool Wear Condition Prediction Using Vibration Signals in High Speed Machining (HSM) of Titanium (Ti-6Al-4V) Alloy”. Procedia Computer Science [online] 50, 270–275. available from <https://www.sciencedirect.com/science/article/pii/S1877050915005505> [13 September 2020]

Kristen + Görmann KG - Werkzeugtechnologien Für Spanabhebende Formgebung - Drehen, Stechen, Fräsen, Hinterlegen, Werkzeuge, Werkzeuge: Willkommen (n.d.) available from <http://www.kristen-goermann.de/68/> [13 September 2020]

Kristoufek, L. (2010) “Rescaled Range Analysis and Detrended FluctuationAnalysis: Finite Sample Properties and ConfidenceIntervals”. AUCO Czech Economic Review 4 (3)

Krizhevsky, A., Sutskever, I., and Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks [online] available from <https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>

Kroha, P. and Škoula, M. (2018) “Hurst Exponent and Trading Signals Derived from Market Time Series”. Proceedings of the 20th International Conference on Enterprise Information Systems [online] available from <https://www.scitepress.org/Papers/2018/66670/66670.pdf> [13 September 2020]

Kwak, G.H., Park, C.W., Lee, K.D., Na, S.I., Ahn, H.Y., and Park, N.W. (2019) Combination of 2D-CNN and Random Forest Models for Crop Classification with UAV Imagery [online] available from <https://a-a-r-s.org/proceeding/ACRS2019/ThP-94.pdf>

Lai, C., Guo, S., Cheng, L., and Wang, W. (2017) “A Comparative Study of Feature Selection Methods for the Discriminative Analysis of Temporal Lobe Epilepsy”. Frontiers in Neurology [online] 8. available from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5770628/> [13 September 2020]

Lai, X., Yang, T., Wang, Z., and Chen, P. (2019) “IoT Implementation of Kalman Filter to Improve Accuracy of Air Quality Monitoring and Prediction”. Applied Sciences [online] 9 (9), 1831. available from <https://www.mdpi.com/2076-3417/9/9/1831/htm> [13 September 2020]

Lambrou, T.P., Anastasiou, C.C., Panayiotou, C.G., and Polycarpou, M.M. (2014) “A Low-Cost Sensor Network for Real-Time Monitoring and Contamination Detection in Drinking Water Distribution Systems”. IEEE Sensors Journal [online] 14 (8), 2765–2772. available from <https://ieeexplore.ieee.org/document/6786370> [13 September 2020]

Learn | OpenEnergyMonitor (n.d.) available from <https://learn.openenergymonitor.org> [13 September 2020]

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998) “Gradient-Based Learning Applied to Document Recognition”. Proceedings of the IEEE 86 (11), 2278–2324

Lee, E.T. and Eun, H.C. (2016) Structural Damage Detection by Power Spectral Density Estimation Using Output-Only Measurement [online] available from <https://www.hindawi.com/journals/sv/2016/8761249/> [13 September 2020]

Lee, S.H., Näppi, J.J., and Yoshida, H. (2012) “Comparative Performance of State-of-the-Art Classifiers in Computer-Aided Detection for CT Colonography”. Lecture Notes in Computer Science 78–87

Lee, S.Y. (2006) In-Process Tool Condition Monitoring Systems in CNC Turning Operations [online] available from <https://lib.dr.iastate.edu/cgi/viewcontent.cgi?referer=https://www.google.co.uk/&httpsredir=1&article=2271&context=rtd> [13 September 2020]

Lee, W.J., Mendis, G.P., and Sutherland, J.W. (2019) “Development of an Intelligent Tool Condition Monitoring System to Identify Manufacturing Tradeoffs and Optimal Machining Conditions”. Procedia Manufacturing [online] 33, 256–263. available from <https://www.sciencedirect.com/science/article/pii/S2351978919305086> [13 September 2020]

Lee, W.K., Ratnam, M.M., and Ahmad, Z.A. (2017) “Detection of Chipping in Ceramic Cutting Inserts from Workpiece Profile during Turning Using Fast Fourier Transform (FFT) and Continuous Wavelet Transform (CWT)”. Precision Engineering 47, 406–423

Leevy, J.L., Khoshgoftaar, T.M., Bauder, R.A., and Seliya, N. (2018) “A Survey on Addressing High-Class Imbalance in Big Data”. Journal of Big Data 5 (1)

Lei, J., Liu, C., and Jiang, D. (2019) “Fault Diagnosis of Wind Turbine Based on Long Short-Term Memory Networks”. Renewable Energy 133, 422–432

Li, H., Du, D., You, X., and He, Q. (2017) Direct Calculation Method of Probability Density from Sampled Vibration Signal Based on Linear Interpolation Method [online] available from <https://www.jvejournals.com/article/18772> [13 September 2020]

Li, H., Wang, X., Liu, C., Zeng, Q., Zheng, Y., Chu, X., Yao, L., Wang, J., Jiao, Y., and Karmakar, C. (2020) “A Fusion Framework Based on Multi-Domain Features and Deep Learning Features of Phonocardiogram for Coronary Artery Disease Detection”. Computers in Biology and Medicine [online] 120, 103733. available from <https://www.sciencedirect.com/science/article/pii/S0010482520301165> [13 September 2020]

Li, K., Wu, Y., Nan, Y., Li, P., and Li, Y. (2017) “Hierarchical Multi-Class Classification in Multimodal Spacecraft Data Using DNN and Weighted Support Vector Machine”. Neurocomputing 259, 55–65

Li, L., Hasegawa, K., and Satoshi T. (2018) Methods and Applications for Modeling and Simulation of Complex Systems 18th Asia Simulation Conference, AsiaSim 2018, Kyoto, Japan, October 27–29, 2018, Proceedings. Singapore Springer Singapore

Li, L., Ota, K., and Dong, M. (2018) “Deep Learning for Smart Industry: Efficient Manufacture Inspection System With Fog Computing”. IEEE Transactions on Industrial Informatics [online] 14 (10), 4665–4673. available from <https://ieeexplore.ieee.org/document/8370640> [26 July 2020]

Li, L., Ota, K., Dong, M. and Borjigin, W. (2017) "Eyes In The Dark: Distributed Scene Understanding For Disaster Management". IEEE Transactions On Parallel And Distributed Systems 28 (12), 3458-3471

Li, S., Liu, G., Tang, X., Lu, J., and Hu, J. (2017) “An Ensemble Deep Convolutional Neural Network Model with Improved D-S Evidence Fusion for Bearing Fault Diagnosis”. Sensors 17 (8), 1729

Li, T., Zhao, Z., Sun, C., Yan, R., and Chen, X. (2020) “Multi-Scale CNN for Multi-Sensor Feature Fusion in Helical Gear Fault Detection”. Procedia Manufacturing [online] 49, 89–93. available from <https://www.sciencedirect.com/science/article/pii/S235197892031653X> [13 September 2020]

Li, W. and Liu, T. (2019) “Time Varying and Condition Adaptive Hidden Markov Model for Tool Wear State Estimation and Remaining Useful Life Prediction in Micro-Milling”. Mechanical Systems and Signal Processing [online] 131, 689–702. available from <https://www.sciencedirect.com/science/article/pii/S0888327019304042#f0030> [19 November 2020]

Li, Y., Du, X., Wan, F., Wang, X., and Yu, H. (2020) “Rotating Machinery Fault Diagnosis Based on Convolutional Neural Network and Infrared Thermal Imaging”. Chinese Journal of Aeronautics [online] 33 (2), 427–438. available from <https://www.sciencedirect.com/science/article/pii/S1000936119303127> [13 September 2020]

Li, Y., Wang, S., Umarov, R., Xie, B., Fan, M., Li, L., and Gao, X. (2018) “DEEPre: Sequence-Based Enzyme EC Number Prediction by Deep Learning”. Bioinformatics [online] 34 (5), 760–769. available from <https://academic.oup.com/bioinformatics/article/34/5/760/4562505> [13 September 2020]

Lin, H., Zeadally, S., Chen, Z., Labiod, H., and Wang, L. (2020) “A Survey on Computation Offloading Modeling for Edge Computing”. Journal of Network and Computer Applications 169, 102781

Lin, J. and Chen, Q. (2014) “A Novel Method for Feature Extraction Using Crossover Characteristics of Nonlinear Data and Its Application to Fault Diagnosis of Rotary Machinery”. Mechanical Systems and Signal Processing [online] 48 (1), 174–187. available from <https://www.sciencedirect.com/science/article/pii/S0888327014001071> [13 September 2020]

Liu, C., Wang, G.F., and Li, Z.M. (2015) “Incremental Learning for Online Tool Condition Monitoring Using Ellipsoid ARTMAP Network Model”. Applied Soft Computing [online] 35, 186–198. available from <https://www.sciencedirect.com/science/article/pii/S1568494615003725> [13 September 2020]

Liu, H., Wang, Y., Li, F., Wang, X., Liu, C., and Pecht, M.G. (2019) “Perceptual Vibration Hashing by Sub-Band Coding: An Edge Computing Method for Condition Monitoring”. IEEE Access [online] 7, 129644–129658. available from <https://ieeexplore.ieee.org/document/8832158> [13 September 2020]

Liu, R., Meng, G., Yang, B., Sun, C., and Chen, X. (2017) “Dislocated Time Series Convolutional Neural Architecture: An Intelligent Fault Diagnosis Approach for Electric Machine”. IEEE Transactions on Industrial Informatics [online] 13 (3), 1310–1320. available from <https://ieeexplore.ieee.org/document/7797508> [13 September 2020]

Liu, Y. and Chen, Y. (2019) “A Normal Distribution Model for Diffuse Radiation versus Incidence Angle”. Solar Energy [online] 186, 60–71. available from <https://www.sciencedirect.com/science/article/pii/S0038092X19304438> [13 September 2020]

Liu, Y., Yan, X., Zhang, C., and Liu, W. (2019) “An Ensemble Convolutional Neural Networks for Bearing Fault Diagnosis Using Multi-Sensor Data”. Sensors (Basel, Switzerland) [online] 19 (23). available from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6929198/> [13 September 2020]

Liu, Z., Du, J., Wang, M., and Ge, S.S. (2020) “ADCM: Attention Dropout Convolutional Module”. Neurocomputing [online] 394, 95–104. available from <https://www.sciencedirect.com/science/article/pii/S092523122030182X#bib0034> [13 September 2020]

Lockridge, G., Dzwonkowski, B., Nelson, R., and Powers, S. (2016) “Development of a Low-Cost Arduino-Based Sonde for Coastal Applications”. Sensors 16 (4), 528

Lotfalinezhad, H. and Maleki, A. (2020) “TTA, a New Approach to Estimate Hurst Exponent with Less Estimation Error and Computational Time”. Physica A: Statistical Mechanics and Its Applications [online] 553, 124093. available from <https://www.sciencedirect.com/science/article/pii/S0378437119322605> [13 September 2020]

Lou, P., Liu, S., Hu, J., Li, R., Xiao, Z., and Yan, J. (2020) “Intelligent Machine Tool Based on Edge-Cloud Collaboration”. IEEE Access 1–1

Low, L.L., Yan, S., Kwan, Y.H., Tan, C.S., and Thumboo, J. (2018) “Assessing the Validity of a Data Driven Segmentation Approach: A 4 Year Longitudinal Study of Healthcare Utilization and Mortality”. PLOS ONE 13 (4), e0195243

Lu, C., Wang, Y., Ragulskis, M., and Cheng, Y. (2016) “Fault Diagnosis for Rotating Machinery: A Method Based on Image Processing”. PLOS ONE 11 (10), e0164111

Lu, S., Lu, Z., and Zhang, Y.D. (2019) “Pathological Brain Detection Based on AlexNet and Transfer Learning”. Journal of Computational Science [online] 30, 41–47. available from <https://www.sciencedirect.com/science/article/pii/S1877750318309116> [13 September 2020]

Lu, S., Shen, S., Huang, J., Dong, M., Lu, J., and Li, W. (2018) “Feature Selection of Laser-Induced Breakdown Spectroscopy Data for Steel Aging Estimation”. Spectrochimica Acta Part B: Atomic Spectroscopy [online] 150, 49–58. available from <https://www.sciencedirect.com/science/article/pii/S058485471830171X#bb0235> [13 September 2020]

Luo, F., Wang, M., Liu, Y., Zhao, X.M., and Li, A. (2019) “DeepPhos: Prediction of Protein Phosphorylation Sites with Deep Learning”. Bioinformatics [online] 35 (16), 2766–2773. available from <https://academic.oup.com/bioinformatics/article/35/16/2766/5270665> [13 September 2020]

Luo, Y. and Huang, Y. (2018) “A New Combined Approach on Hurst Exponent Estimate and Its Applications in Realized Volatility”. Physica A: Statistical Mechanics and Its Applications [online] 492, 1364–1372. available from <https://www.sciencedirect.com/science/article/pii/S0378437117311524> [13 September 2020]

Madeinchina2025 (2020) available from <http://english.www.gov.cn/2016special/madeinchina2025/> [13 September 2020]

Madhusudana, C., Budati, S., Gangadhar, N., Kumar, H., and Narendranath, S. (2016) “Fault Diagnosis Studies of Face Milling Cutter Using Machine Learning Approach”. Journal of Low Frequency Noise, Vibration and Active Control 35 (2), 128–138

Madhusudana, C.K., Kumar, H., and Narendranath, S. (2019) “Vibration-Based Fault Diagnosis of a Face Milling Tool Using Empirical Mode Decomposition Features and Artificial Neural Network”. International Journal of Condition Monitoring [online] 9 (2), 25–34. available from <https://www.ingentaconnect.com/content/bindt/ijcm/2019/00000009/00000002/art00001> [13 September 2020]

Mahecha, M.D., Martínez, A., Lischeid, G., and Beck, E. (2007) “Nonlinear Dimensionality Reduction: Alternative Ordination Approaches for Extracting and Visualizing Biodiversity Patterns in Tropical Montane Forest Vegetation Data”. Ecological Informatics 2 (2), 138–149

Mali, R., Telsang, M.T., and Gupta, T.V.K. (2017) “Real Time Tool Wear Condition Monitoring in Hard Turning of Inconel 718 Using Sensor Fusion System”. Materials Today: Proceedings 4 (8), 8605–8612

Mandal, S. (2014) "Applicability Of Tool Condition Monitoring Methods Used For Conventional Milling In Micromilling: A Comparative Review". Journal Of Industrial Engineering 2014, 1-8

Mao, W., He, J., Tang, J., and Li, Y. (2018) “Predicting Remaining Useful Life of Rolling Bearings Based on Deep Feature Representation and Long Short-Term Memory Neural Network”. Advances in Mechanical Engineering 10 (12), 168781401881718

Mao, Y. and Yang, Y. (2019) A Wrapper Feature Subset Selection Method Based on Randomized Search and Multilayer Structure [online] available from <https://www.hindawi.com/journals/bmri/2019/9864213/> [13 September 2020]

Mares, M.A., Wang, S., and Guo, and Y. (2016) Combining Multiple Feature Selection Methods and Deep Learning for High-Dimensional Data [online] available from <http://www.ibai-publishing.org/journal/issue_mldm/2016_july/mldm_9_1_27_45.php>

Martínez-Arellano, G., Terrazas, G., and Ratchev, S. (2019) “Tool Wear Classification Using Time Series Imaging and Deep Learning”. The International Journal of Advanced Manufacturing Technology [online] 104 (9), 3647–3662. available from <https://link.springer.com/article/10.1007%2Fs00170-019-04090-6> [13 September 2020]

Marusina, M.Y., Mochalina, A.P., Frolova, E.P., Satikov, V.I., Barchuk, A.A., Kuznetcov, V.I., Gaidukov, V.S., and Tarakanov, S.A. (2017) “MRI Image Processing Based on Fractal Analysis”. Asian Pacific Journal of Cancer Prevention 18 (1)

Maurya, S., Singh, V., and Verma, N.K. (2020) “Condition Monitoring of Machines Using Fused Features From EMD-Based Local Energy With DNN”. IEEE Sensors Journal [online] 20 (15), 8316–8327. available from <https://ieeexplore.ieee.org/document/8758425> [13 September 2020]

McParland, D., Baron, S., O’Rourke, S., Dowling, D., Ahearne, E. and Parnell, A. (2017) "Prediction Of Tool-Wear In Turning Of Medical Grade Cobalt Chromium Molybdenum Alloy (ASTM F75) Using Non-Parametric Bayesian Models". Journal Of Intelligent Manufacturing 30 (3), 1259-1270

Mehta, P., Bukov, M., Wang, C.H., Day, A.G.R., Richardson, C., Fisher, C.K., and Schwab, D.J. (2019) “A High-Bias, Low-Variance Introduction to Machine Learning for Physicists”. Physics Reports [online] 810, 1–124. available from <https://www.sciencedirect.com/science/article/pii/S0370157319300766> [7 December 2019]

Meng, C., Yi, X., Su, L., Gao, J., and Zheng, Y. (2017) “City-Wide Traffic Volume Inference with Loop Detector Data and Taxi Trajectories”. Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems

Mikolov, T. and Zweig, G. (2012) Context Dependent Recurrent Neural Network Language Model [online] available from <https://ieeexplore.ieee.org/document/6424228> [13 September 2020]

Minbashi, N., Bagheri, M., Golroo, A., and Arasteh Khouy, I. (2016) “Use of Power Spectral Density for Effective Turnout Geometry Maintenance”. Transportation Research Record: Journal of the Transportation Research Board 2545 (1), 79–88

Mitiche, I., Morison, G., Nesbitt, A., Hughes-Narborough, M., Boreham, P., and Stewart, B.G. (2017) An Evaluation of Total Variation Signal Denoising Methods for Partial Discharge Signals [online] available from <https://pureportal.strath.ac.uk/en/publications/an-evaluation-of-total-variation-signal-denoising-methods-for-par> [13 September 2020]

Mohan, N. and Kangasharju, J. (2016) “Edge-Fog Cloud: A Distributed Cloud for Internet of Things Computations”. 2016 Cloudification of the Internet of Things (CIoT) [online] available from <https://ieeexplore.ieee.org/abstract/document/7872914/> [6 December 2019]

Mohanraj, T., Shankar, S., Rajasekar, R., Deivasigamani, R., and Arunkumar, P.M. (2019) “Tool Condition Monitoring in the Milling Process with Vegetable Based Cutting Fluids Using Vibration Signatures”. Materials Testing 61 (3), 282–288

Mohanty, S., Gupta, K.K., and Raju, K.S. (2018) “Hurst Based Vibro-Acoustic Feature Extraction of Bearing Using EMD and VMD”. Measurement 117, 200–220

Motlagh, N., Mohammadrezaei, M., Hunt, J. and Zakeri, B. (2020) "Internet Of Things (Iot) And The Energy Sector". Energies 13 (2), 494

Mourtzis, D., Vlachou, E., Milas, N., and Xanthopoulos, N. (2016) “A Cloud-Based Approach for Maintenance of Machine Tools and Equipment Based on Shop-Floor Monitoring”. Procedia CIRP 41, 655–660

Mukhopadhyay, C.K., Jayakumar, T., Raj, B., and Venugopal, S. (2012) “Statistical Analysis of Acoustic Emission Signals Generated during Turning of a Metal Matrix Composite”. Journal of the Brazilian Society of Mechanical Sciences and Engineering 34 (2), 145–154

Nagy, J., Oláh, J., Erdei, E., Máté, D. and Popp, J. (2018) "The Role And Impact Of Industry 4.0 And The Internet Of Things On The Business Strategy Of The Value Chain—The Case Of Hungary". Sustainability 10 (10), 3491

Nair, V. and Hinton, G. (2010) Rectified Linear Units Improve Restricted Boltzmann Machines [online] available from <https://www.cs.toronto.edu/~fritz/absps/reluICML.pdf>

Nalepa, J. and Kawulok, M. (2018) “Selecting Training Sets for Support Vector Machines: A Review”. Artificial Intelligence Review 52 (2), 857–900

Nanda, M.A., Seminar, K.B., Nandika, D., and Maddu, A. (2018) “A Comparison Study of Kernel Functions in the Support Vector Machine and Its Application for Termite Detection”. Information [online] 9 (1), 5. available from <https://www.mdpi.com/2078-2489/9/1/5> [13 September 2020]

Nastac, S. (2018) On Vibration Joint Time-Frequency Investigations of CNC Milling Machines for Tool Trajectory Task Conformity Estimation [online] available from <https://www.hindawi.com/journals/sv/2018/7375057/> [13 September 2020]

Neef, B., Bartels, J., and Thiede, S. (2018) Tool Wear and Surface Quality Monitoring Using High Frequency CNC Machine Tool Current Signature [online] available from <https://ieeexplore.ieee.org/document/8472037> [13 September 2020]

Nguyen Gia, T., Queralta, J.P., and Westerlund, T. (2020) “Exploiting LoRa, Edge, and Fog Computing for Traffic Monitoring in Smart Cities”. LPWAN Technologies for IoT and M2M Applications 347–371

Ning, H., Li, Y., Shi, F., and Yang, L.T. (2020) “Heterogeneous Edge Computing Open Platforms and Tools for Internet of Things”. Future Generation Computer Systems [online] 106, 67–76. available from <https://www.sciencedirect.com/science/article/pii/S0167739X19312841> [13 September 2020]

Niu, B., Sun, J., and Yang, B. (2020) “Multisensory Based Tool Wear Monitoring for Practical Applications in Milling of Titanium Alloy”. Materials Today: Proceedings [online] 22, 1209–1217. available from <https://www.sciencedirect.com/science/article/pii/S2214785319341379> [13 September 2020]

Núñez, J.C., Cabido, R., Pantrigo, J.J., Montemayor, A.S., and Vélez, J.F. (2018) “Convolutional Neural Networks and Long Short-Term Memory for Skeleton-Based Human Activity and Hand Gesture Recognition”. Pattern Recognition [online] 76, 80–94. available from <https://www.sciencedirect.com/science/article/pii/S0031320317304405>

Nwankpa, C.E., Ijomah, W., Gachagan, A., and Marshall, S. (2018) Activation Functions: Comparison of Trends in Practice and Research for Deep Learning [online] available from <https://arxiv.org/pdf/1811.03378.pdf>

NXP Semiconductors | Automotive, Security, IoT (2019) available from <https://www.nxp.com/> [28 November 2019]

O’Grady, M.J., Langton, D., and O’Hare, G.M.P. (2019) “Edge Computing: A Tractable Model for Smart Agriculture?” Artificial Intelligence in Agriculture

Opałka, S., Stasiak, B., Szajerman, D., and Wojciechowski, A. (2018) “Multi-Channel Convolutional Neural Networks Architecture Feeding for Effective EEG Mental Tasks Classification”. Sensors 18 (10), 3451

Oral, E. and Unal, G. (2019) “Modeling and Forecasting Time Series of Precious Metals: A New Approach to Multifractal Data”. Financial Innovation 5 (1)

Painuli, S., Elangovan, M., and Sugumaran, V. (2014) “Tool Condition Monitoring Using K-Star Algorithm”. Expert Systems with Applications [online] 41 (6), 2638–2643. available from <https://www.sciencedirect.com/science/article/pii/S0957417413009056> [13 September 2020]

Pandiyan, V., Caesarendra, W., Tjahjowidodo, T., and Tan, H.H. (2018) “In-Process Tool Condition Monitoring in Compliant Abrasive Belt Grinding Process Using Support Vector Machine and Genetic Algorithm”. Journal of Manufacturing Processes [online] 31, 199–213. available from <https://www.sciencedirect.com/science/article/pii/S1526612517303481> [13 September 2020]

Pappachan, B.K., Caesarendra, W., Tjahjowidodo, T., and Wijaya, T. (2017) “Frequency-domain Analysis of Sensor Data for Event Classification in Real-Time Robot Assisted Deburring”. Sensors [online] 17 (6), 1247. available from <https://www.mdpi.com/1424-8220/17/6/1247> [13 September 2020]

Ortiz, J.P., Aguilera, R.C., Balankin, A.S., Ortiz, M.P., Rodriguez, J.C.T., Mosqueda, M.A.A., Cruz, M.A.M., and Yu, W. (2016) “Seismic Activity Seen Through Evolution of the Hurst Exponent Model in 3D”. Fractals 24 (04), 1650045

Patra, K., Jha, A.K., Szalay, T., Ranjan, J., and Monostori, L. (2017) “Artificial Neural Network Based Tool Condition Monitoring in Micro Mechanical Peck Drilling Using Thrust Force Signals”. Precision Engineering 48, 279–291

Pes, B. (2019) “Ensemble Feature Selection for High-Dimensional Data: A Stability Analysis across Multiple Domains”. Neural Computing and Applications 32 (10), 5951–5973

Pham, M., Mengistu, Y., Do, H., and Sheng, W. (2018) “Delivering Home Healthcare through a Cloud-Based Smart Home Environment (CoSHE)”. Future Generation Computer Systems [online] 81, 129–140. available from <https://www.sciencedirect.com/science/article/pii/S0167739X17302194> [18 November 2019]

Poernomo, A. and Kang, D.K. (2018) “Biased Dropout and Crossmap Dropout: Learning towards Effective Dropout Regularization in Convolutional Neural Network”. Neural Networks [online] 104, 60–67. available from <https://www.sciencedirect.com/science/article/pii/S0893608018301096> [18 April 2019]

Pollreisz, D. and TaheriNejad, N. (2019) “Detection and Removal of Motion Artifacts in PPG Signals”. Mobile Networks and Applications

Qian, Y., Yan, R., and Hu, S. (2014) “Bearing Degradation Evaluation Using Recurrence Quantification Analysis and Kalman Filter”. IEEE Transactions on Instrumentation and Measurement 63 (11), 2599–2610

Qiao, H., Wang, T., and Wang, P. (2020) “A Tool Wear Monitoring and Prediction System Based on Multiscale Deep Learning Models and Fog Computing”. The International Journal of Advanced Manufacturing Technology 108 (7–8), 2367–2384

Qin, Y., Mo, L., Guo, W., and Li, F. (n.d.) Combination of 3D CNNs and LSTMs and Its Application In Activity Recognition [online] available from <http://en.cnki.com.cn/Article_en/CJFDTotal-IKJS201702007.htm> [13 September 2020]

Qin, Y., Wang, H., Huang, Y., and Tang, X. (2019) Real-Time Road Profile Identification and Monitoring: Theory and Application [online] Morgan & Claypool Publishers. available from <https://books.google.co.uk/books?id=reiEDwAAQBAJ&dq=dimension+feature+and+dimensionless+feature+all+belongs+to+statistic&source=gbs_navlinks_s> [13 September 2020]

Ostasevicius, V., Markevicius, V., Jurenas, V., Zilys, M., Cepenas, M., Kizauskiene, L., and Gyliene, V. (2015) “Cutting Tool Vibration Energy Harvesting for Wireless Sensors Applications”. Sensors and Actuators A: Physical [online] 233, 310–318. available from <https://www.sciencedirect.com/science/article/pii/S0924424715300686> [15 September 2020]

Queralta, J.P., Gia, T.N., Tenhunen, H., and Westerlund, T. (2019) Edge-AI in LoRa-Based Health Monitoring: Fall Detection System with Fog Computing and LSTM Recurrent Neural Networks [online] available from <https://ieeexplore.ieee.org/abstract/document/8768883> [13 September 2020]

Prajwala, T.R. (2015) “A Comparative Study on Decision Tree and Random Forest Using R Tool”. IJARCCE 196–199

Rad, J.S., Zhang, Y., Aghazadeh, F., and Chen, Z.C. (2014) A Study on Tool Wear Monitoring Using Time-Frequency Transformation Techniques [online] available from <https://ieeexplore.ieee.org/document/6912718> [13 September 2020]

Rafael, R., Shirley, A. J. and Liveris, A. (2014) Report To The President Accelerating U.S. Advanced Manufacturing. Report, the President's Council of Advisors on Science and Technology Saaty, T. L. (1977) 'A scaling method for priorities in hierarchical structures', Journal of Mathematical Psychology, 15(3), pp. 234–281.

Rahman, Md.M., Islam, Md.S., Sassi, R., and Aktaruzzaman, Md. (2019) “Convolutional Neural Networks Performance Comparison for Handwritten Bengali Numerals Recognition”. SN Applied Sciences 1 (12)

Rajamani, D., Esakki, B., Arunkumar, P. and Velu, R. (2018) "Fuzzy Logic-Based Expert System For Prediction Of Wear Rate In Selective Inhibition Sintered HDPE Parts". Materials Today: Proceedings 5 (2), 6072-6081

Ramachandran, P., Zoph, B., and Le, Q.V. (2017) Searching for Activation Functions [online] available from <https://arxiv.org/abs/1710.05941> [1 December 2019]

Ramos-Requena, J.P., Trinidad-Segovia, J.E., and Sánchez-Granero, M.A. (2017) “Introducing Hurst Exponent in Pair Trading”. Physica A: Statistical Mechanics and Its Applications [online] 488, 39–45. available from <https://www.sciencedirect.com/science/article/pii/S0378437117306738> [13 September 2020]

Rasheed, I., Zhang, L., and Hu, F. (2020) “A Privacy Preserving Scheme for Vehicle-to-Everything Communications Using 5G Mobile Edge Computing”. Computer Networks 176, 107283

Raspberry Pi 3 Model B+ - Raspberry Pi (2014) available from <https://www.raspberrypi.org/products/raspberry-pi-3-model-b-plus/> [13 March 2019]

Rehman, M.H. ur, Jayaraman, P.P., Malik, S.U.R., Khan, A.U.R., and Gaber, M.M. (2017) “RedEdge: A Novel Architecture for Big Data Processing in Mobile Edge Computing Environments”. Journal of Sensor and Actuator Networks [online] 6 (3), 17. available from <https://www.mdpi.com/2224-2708/6/3/17> [13 September 2020]

Richmond, D.L., Kainmüller, D., Yang, M., Myers, E., and Rother, C. (2015) Mapping Stacked Decision Forests to Deep and Sparse Convolutional Neural Networks for Semantic Segmentation [online] available from <https://www.semanticscholar.org/paper/Mapping-Stacked-Decision-Forests-to-Deep-and-Sparse-Richmond-Kainm%C3%BCller/4568d3ab1304f0da093b4354c998cd44c8300e93> [13 September 2020]

Rizal, M., Ghani, J.A., Nuawi, M.Z., and Haron, C.H.C. (2017) “Cutting Tool Wear Classification and Detection Using Multi-Sensor Signals and Mahalanobis-Taguchi System”. Wear [online] 376–377, 1759–1765. available from <https://www.sciencedirect.com/science/article/pii/S0043164817303095#bib5> [13 September 2020]

Rodriguez, D., Herraiz, I., Harrison, R., Dolado, J., and Riquelme, J.C. (2014) “Preliminary Comparison of Techniques for Dealing with Imbalance in Software Defect Prediction”. Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering - EASE ’14

Roheda, S., Krim, H., Luo, Z.Q., and Wu, T. (2018) Decision Level Fusion: An Event Driven Approach [online] available from <https://ieeexplore.ieee.org/document/8553412> [13 September 2020]

Roy, A. and Todorovic, S. (2016) Monocular Depth Estimation Using Neural Regression Forest [online] available from <http://web.engr.oregonstate.edu/~sinisa/research/publications/cvpr16_NRF.pdf> [13 September 2020]

Rusinek, R. and Borowiec, M. (2015) “Stability Analysis of Titanium Alloy Milling by Multiscale Entropy and Hurst Exponent”. The European Physical Journal Plus 130 (10)

Sahu, S.N. and Nayak, N.C. (2018) “Multi-Criteria Decision Making with PCA in EDM of A2 Tool Steel”. Materials Today: Proceedings 5 (9), 18641–18648

Sainath, T.N., Vinyals, O., Senior, A., and Sak, H. (2015) Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks [online] available from <https://ieeexplore.ieee.org/document/7178838> [13 September 2020]

Saini, S., Ahuja, I.S., and Sharma, V.S. (2012) “Influence of Cutting Parameters on Tool Wear and Surface Roughness in Hard Turning of AISI H11 Tool Steel Using Ceramic Tools”. International Journal of Precision Engineering and Manufacturing 13 (8), 1295–1302

Sandvik Groove or Slot Milling available from <https://www.sandvik.coromant.com/en-us/knowledge/milling/application_overview/slot_milling/end_milling_of_slots> [13 September 2017]

Sarlis, N.V., Skordas, E.S., Mintzelas, A., and Papadopoulou, K.A. (2018) “Micro-Scale, Mid-Scale, and Macro-Scale in Global Seismicity Identified by Empirical Mode Decomposition and Their Multifractal Characteristics”. Scientific Reports [online] 8. available from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6003985/> [13 September 2020]

Satija, U., Ramkumar, B., and Sabarimalai Manikandan, M. (2017) “Real-Time Signal Quality-Aware ECG Telemetry System for IoT-Based Health Care Monitoring”. IEEE Internet of Things Journal 4 (3), 815–823

Schmidt, J., Marques, M., Botti, S. and Marques, M. (2019) "Recent Advances And Applications Of Machine Learning In Solid-State Materials Science". Npj Computational Materials 5 (1)

Setz, B. and Aiello, M. (2020) “Towards Real-Time Monitoring of Data Centers Using Edge Computing”. Service-Oriented and Cloud Computing 141–148

Shankar, S., Mohanraj, T. and Pramanik, A. (2019) "Tool Condition Monitoring While Using Vegetable Based Cutting Fluids During Milling Of Inconel 625". Journal Of Advanced Manufacturing Systems 18 (04), 563-581

Shanthi, T. and Sabeenian, R.S. (2019) “Modified Alexnet Architecture for Classification of Diabetic Retinopathy Images”. Computers & Electrical Engineering 76, 56–64

Sharaff, A. and Gupta, H. (2019) “Extra-Tree Classifier with Metaheuristics Approach for Email Classification”. Advances in Intelligent Systems and Computing [online] 189–197. available from <https://link.springer.com/chapter/10.1007%2F978-981-13-6861-5_17> [13 September 2020]

Shawky, O.A., Hagag, A., El-Dahshan, E.S.A., and Ismail, M.A. (2020) “Remote Sensing Image Scene Classification Using CNN-MLP with Data Augmentation”. Optik [online] 221, 165356. available from <https://www.sciencedirect.com/science/article/pii/S003040262031192X> [13 September 2020]

Sheng, H. (2012) Model-based Tool Condition Monitoring for Ball-nose End Milling. Huazhong University of Science and Technology

Shi, C., Panoutsos, G., Luo, B., Liu, H., Li, B., and Lin, X. (2019) “Using Multiple-Feature-Spaces-Based Deep Learning for Tool Condition Monitoring in Ultraprecision Manufacturing”. IEEE Transactions on Industrial Electronics [online] 66 (5), 3794–3803. available from <https://ieeexplore.ieee.org/document/8421243> [13 September 2020]

Simon, G.D. and Deivanathan, R. (2019) “Early Detection of Drilling Tool Wear by Vibration Data Acquisition and Classification”. Manufacturing Letters [online] 21, 60–65. available from <https://www.sciencedirect.com/science/article/pii/S2213846319300410> [8 January 2020]

Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition [online] available from <https://arxiv.org/abs/1409.1556>

Song, S., Lan, C., xing, J., Zeng, W., and Liu, J. (2017) “An End-to-End Spatio-Temporal Attention Model for Human Action Recognition from Skeleton Data”. in AAAI’17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. held February 2017

Soudani, A. and Barhoumi, W. (2019) “An Image-Based Segmentation Recommender Using Crowdsourcing and Transfer Learning for Skin Lesion Extraction”. Expert Systems with Applications [online] 118, 400–410. available from <https://www.sciencedirect.com/science/article/pii/S0957417418306791> [13 September 2020]

Stollenga, M.F., Byeon, W., Liwicki, M., and Schmidhuber, J. (2015) Parallel Multi-Dimensional LSTM, With Application to Fast Biomedical Volumetric Image Segmentation [online] available from <http://papers.nips.cc/paper/5642-parallel-multi-dimensional-lstm-with-application-to-fast-biomedical-volumetric-image-segmentation> [13 September 2020]

Sun, G., Qin, A., Zhang, Q., Hu, Q., and Si, X. (2013) A Compound Fault Integrated Diagnosis Method for Rotating Machinery Base on Dimensionless Immune Detector [online] available from <https://ieeexplore.ieee.org/document/6561724> [13 September 2020]

Sun, H., Zhang, J., Mo, R., and Zhang, X. (2020) “In-Process Tool Condition Forecasting Based on a Deep Learning Method”. Robotics and Computer-Integrated Manufacturing [online] 64, 101924. available from <https://www.sciencedirect.com/science/article/pii/S0736584519303655#bib0031> [13 September 2020]

Sun, L.L. and Zhao, L. (2011) “Research of Tool Wear Monitoring Based on Hurst Exponent Extraction of Cutting Surface Texture”. Advanced Materials Research 411, 163–166

Sun, S., Hu, X., Cai, W., and Zhong, J. (2019) “Tool Breakage Detection of Milling Cutter Insert Based on SVM”. IFAC-PapersOnLine [online] 52 (13), 1549–1554. available from <https://www.sciencedirect.com/science/article/pii/S2405896319314016> [13 September 2020]

Sun, W., Yao, B., Zeng, N., Chen, B., He, Y., Cao, X., and He, W. (2017) “An Intelligent Gear Fault Diagnosis Methodology Using a Complex Wavelet Enhanced Convolutional Neural Network”. Materials [online] 10 (7), 790. available from <https://www.mdpi.com/1996-1944/10/7/790/htm> [13 September 2020]

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2014) Going Deeper with Convolutions [online] available from <https://arxiv.org/abs/1409.4842> [29 July 2019]

Tabassam, S. (2017) “Security and Privacy Issues in Cloud Computing Environment”. Journal of Information Technology & Software Engineering 07 (05)

Tariq, N., Asim, M., Al-Obeidat, F., Zubair Farooqi, M., Baker, T., Hammoudeh, M., and Ghafir, I. (2019) “The Security of Big Data in Fog-Enabled IoT Applications Including Blockchain: A Survey”. Sensors 19 (8), 1788

Terrazas, G., Martínez-Arellano, G., Benardos, P., and Ratchev, S. (2018) “Online Tool Wear Classification during Dry Machining Using Real Time Cutting Force Measurements and a CNN Approach”. Journal of Manufacturing and Materials Processing 2 (4), 72

Terrazas, G., Martínez-Arellano, G., Benardos, P., and Ratchev, S. (2018) “Online Tool Wear Classification during Dry Machining Using Real Time Cutting Force Measurements and a CNN Approach”. Journal of Manufacturing and Materials Processing 2 (4), 72

Tra, V., Khan, S.A., and Kim, J.M. (2018) “Diagnosis of Bearing Defects under Variable Speed Conditions Using Energy Distribution Maps of Acoustic Emission Spectra and Convolutional Neural Networks”. The Journal of the Acoustical Society of America 144 (4), EL322–EL327

Tseng, C.W., Tseng, F.H., Yang, Y.T., Liu, C.C., and Chou, L.D. (2018) Task Scheduling for Edge Computing with Agile VNFs On-Demand Service Model toward 5G and Beyond [online] available from <https://www.hindawi.com/journals/wcmc/2018/7802797/> [13 September 2020]

Uddeholm Corrax CUTTING DATA RECOMMENDATIONS (2007) available from <https://www.uddeholm.com/files/Cutting_Data_Corrax_eng.pdf>

Uddin, Md.Z. (2019) “A Wearable Sensor-Based Activity Prediction System to Facilitate Edge Computing in Smart Healthcare System”. Journal of Parallel and Distributed Computing [online] 123, 46–53. available from <https://www.sciencedirect.com/science/article/pii/S0743731518306270#sec2> [13 September 2020]

Ullah, A., Ahmad, J., Muhammad, K., Sajjad, M., and Baik, S.W. (2018) “Action Recognition in Video Sequences Using Deep Bi-Directional LSTM With CNN Features”. IEEE Access [online] 6, 1155–1166. available from <https://ieeexplore.ieee.org/document/8121994> [13 September 2020]

Ullah, I., Qian, S., Deng, Z., and Lee, J.H. (2020) “Extended Kalman Filter-Based Localization Algorithm by Edge Computing in Wireless Sensor Networks”. Digital Communications and Networks

Vazquez, K.P.M., Giardini, C., and Ceretti, E. (2014) “Cutting Force Modeling”. CIRP Encyclopedia of Production Engineering [online] 315–329. available from <https://link.springer.com/referenceworkentry/10.1007%2F978-3-642-20617-7_6399> [13 September 2020]

Vela-Martínez, L., Jáuregui-Correa, J.C., and Álvarez-Ramírez, J. (2009) “Characterization of Machining Chattering Dynamics: An R/S Scaling Analysis Approach”. International Journal of Machine Tools and Manufacture [online] 49 (11), 832–842. available from <https://www.sciencedirect.com/science/article/pii/S0890695509000935> [13 September 2020]

Verderame, L., Merelli, I., Morganti, L., Corni, E., Cesini, D., D’Agostino, D., and Merlo, A. (2019) “A Secure Cloud-Edges Computing Architecture for Metagenomics Analysis”. Future Generation Computer Systems

Villalonga, A., Castaño, F., Beruvides, G., Haber, R., Strzelczak, S., and Kossakowska, J. (2019) Visual Analytics Framework for Condition Monitoring in Cyber-Physical Systems [online] available from <https://ieeexplore.ieee.org/document/8885611>

Walia, S. (2018) Low Power Semiconductor Devices and Processes for Emerging Applications in Communications, Computing, and Sensing [online] CRC Press. available from <https://books.google.co.uk/books?id=wjBtDwAAQBAJ&dq=Kalman+gain+will+iterates+to+a+stable+value&source=gbs_navlinks_s> [13 September 2020]

Wan, J., Chen, B., Wang, S., Xia, M., Li, D., and Liu, C. (2018) “Fog Computing for Energy-Aware Load Balancing and Scheduling in Smart Factory”. IEEE Transactions on Industrial Informatics [online] 14 (10), 4548–4556. available from <https://ieeexplore.ieee.org/abstract/document/8323243> [13 September 2020]

Wang, B., Ning, L., and Kong, Y. (2019) “Integration of Unsupervised and Supervised Machine Learning Algorithms for Credit Risk Assessment”. Expert Systems with Applications [online] 128, 301–315. available from <https://www.sciencedirect.com/science/article/pii/S0957417419301472> [13 September 2020]

Wang, G., Yang, Y., and Guo, Z. (2013) “Hybrid Learning Based Gaussian ARTMAP Network for Tool Condition Monitoring Using Selected Force Harmonic Features”. Sensors and Actuators A: Physical [online] 203, 394–404. available from <https://www.sciencedirect.com/science/article/pii/S092442471300472X#bib0075> [13 September 2020]

Wang, G., Yang, Y., and Li, Z. (2014) “Force Sensor Based Tool Condition Monitoring Using a Heterogeneous Ensemble Learning Model”. Sensors 14 (11), 21588–21602

Wang, H., Li, S., Song, L., and Cui, L. (2019) “A Novel Convolutional Neural Network Based Fault Recognition Method via Image Fusion of Multi-Vibration-Signals”. Computers in Industry [online] 105, 182–190. available from <https://www.sciencedirect.com/science/article/pii/S0166361518305190#fig0020> [13 September 2020]

Wang, J., Wang, P., and Gao, R.X. (2015) “Enhanced Particle Filter for Tool Wear Prediction”. Journal of Manufacturing Systems [online] 36, 35–45. available from <https://www.sciencedirect.com/science/article/pii/S0278612515000217> [13 September 2020]

Wang, J., Xie, J., Zhao, R., Mao, K., and Zhang, L. (2016) “A New Probabilistic Kernel Factor Analysis for Multisensory Data Fusion: Application to Tool Condition Monitoring”. IEEE Transactions on Instrumentation and Measurement [online] 65 (11), 2527–2537. available from <https://ieeexplore.ieee.org/document/7517325> [13 September 2020]

Wang, P., Zhang, X., and Hao, Y. (2019) A Method Combining CNN and ELM for Feature Extraction and Classification of SAR Image [online] available from <https://www.hindawi.com/journals/js/2019/6134610/> [13 September 2020]

Wang, R., Peng, C., Gao, J., Gao, Z., and Jiang, H. (2019) “A Dilated Convolution Network-Based LSTM Model for Multi-Step Prediction of Chaotic Time Series”. Computational and Applied Mathematics 39 (1)

Wang, W.F., Qiu, X.H., Chen, C.S., Lin, B., and Zhang, H.M. (2018) Application Research On Long Short-Term Memory Network In Fault Diagnosis [online] available from <https://ieeexplore.ieee.org/abstract/document/8527031> [13 September 2020]

Wang, X., Zheng, Y., Zhao, Z., and Wang, J. (2015) “Bearing Fault Diagnosis Based on Statistical Locally Linear Embedding”. Sensors 15 (7), 16225–16247

Wang, Y., Peng, Y., Zi, Y., Jin, X., and Tsui, K.L. (2016) “A Two-Stage Data-Driven-Based Prognostic Approach for Bearing Degradation Problem”. IEEE Transactions on Industrial Informatics 12 (3), 924–932

Wen, L., Li, X., Gao, L., and Zhang, Y. (2018) “A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method”. IEEE Transactions on Industrial Electronics [online] 65 (7), 5990–5998. available from <https://ieeexplore.ieee.org/abstract/document/8114247> [13 September 2020]

Wu, C.H., Guo, Y., and Fan, Y.J. (2012) “Gear Vibration Monitoring System Based on Virtual Instruments”. Applied Mechanics and Materials 187, 161–164

Wu, D., Liu, S., Zhang, L., Terpenny, J., Gao, R.X., Kurfess, T., and Guzzo, J.A. (2017) “A Fog Computing-Based Framework for Process Monitoring and Prognosis in Cyber-Manufacturing”. Journal of Manufacturing Systems 43, 25–34

Wu, J., Hu, K., Cheng, Y., Zhu, H., Shao, X., and Wang, Y. (2020) “Data-Driven Remaining Useful Life Prediction via Multiple Sensor Signals and Deep Long Short-Term Memory Neural Network”. ISA Transactions [online] 97, 241–250. available from <https://www.sciencedirect.com/science/article/pii/S0019057819302939> [13 September 2020]

Wu, J., Su, Y., Cheng, Y., Shao, X., Deng, C., and Liu, C. (2018) “Multi-Sensor Information Fusion for Remaining Useful Life Prediction of Machining Tools by Adaptive Network Based Fuzzy Inference System”. Applied Soft Computing [online] 68, 13–23. available from <https://www.sciencedirect.com/science/article/pii/S1568494618301741> [13 September 2020]

Wu, W. and Faisal, S. (2019) “A Data-Driven Principal Component Analysis-Support Vector Machine Approach for Breast Cancer Diagnosis: Comparison and Application”. Transactions of the Institute of Measurement and Control 42 (7), 1301–1312

Wu, X., Dunne, R., Zhang, Q., and Shi, W. (2017) “Edge Computing Enabled Smart Firefighting”. Proceedings of the Fifth ACM/IEEE Workshop on Hot Topics in Web Systems and Technologies - HotWeb ’17

Xie, Z., Li, J., and Lu, Y. (2018) “Feature Selection and a Method to Improve the Performance of Tool Condition Monitoring”. The International Journal of Advanced Manufacturing Technology 100 (9–12), 3197–3206

Xu, B., Wang, W., Falzon, G., Kwan, P., Guo, L., Chen, G., Tait, A., and Schneider, D. (2020) “Automated Cattle Counting Using Mask R-CNN in Quadcopter Vision System”. Computers and Electronics in Agriculture [online] 171, 105300. available from <https://www.sciencedirect.com/science/article/pii/S0168169919320149> [13 September 2020]

Xu, C., Bao, J., Liu, P., and Wang, W. (2017) “Investigation of Contributing Factors to Extremely Severe Traffic Crashes Using Survival Theory”. International Journal of Injury Control and Safety Promotion 25 (2), 141–153

Xu, C., Dou, J., Chai, Y., Li, H., Shi, Z. and Xu, J. (2018) "The Relationships Between Cutting Parameters, Tool Wear, Cutting Force And Vibration". Advances In Mechanical Engineering 10 (1), 168781401775043

Xu, L., Pennacchi, P., and Chatterton, S. (2020) “A New Method for the Estimation of Bearing Health State and Remaining Useful Life Based on the Moving Average Cross-Correlation of Power Spectral Density”. Mechanical Systems and Signal Processing [online] 139, 106617. available from <https://www.sciencedirect.com/science/article/pii/S0888327020300030> [13 September 2020]

Yamashita, R., Nishio, M., Do, R.K.G., and Togashi, K. (2018) “Convolutional Neural Networks: An Overview and Application in Radiology”. Insights into Imaging 9 (4), 611–629

Yang, H., Zhang, J., Chen, L., Zhang, H., and Liu, S. (2019) Fault Diagnosis of Reciprocating Compressor Based on Convolutional Neural Networks with Multisource Raw Vibration Signals [online] available from <https://www.hindawi.com/journals/mpe/2019/6921975/> [13 September 2020]

Yang, W.A., Zhou, Q., and Tsui, K.L. (2015) “Differential Evolution-Based Feature Selection and Parameter Optimisation for Extreme Learning Machine in Tool Wear Estimation”. International Journal of Production Research 54 (15), 4703–4721

Yang, Y., Fang, Q., and Shen, H.B. (2019) “Predicting Gene Regulatory Interactions Based on Spatial Gene Expression Data and Deep Learning”. PLOS Computational Biology 15 (9), e1007324

Yang, Y., Guo, Y., Huang, Z., Chen, N., Li, L., Jiang, Y., and He, N. (2019) “Research on the Milling Tool Wear and Life Prediction by Establishing an Integrated Predictive Model”. Measurement [online] 145, 178–189. available from <https://www.sciencedirect.com/science/article/pii/S0263224119304245> [13 September 2020]

Yao, Y., Wang, H., Li, S., Liu, Z., Gui, G., Dan, Y., and Hu, J. (2018) “End-To-End Convolutional Neural Network Model for Gear Fault Diagnosis Based on Sound Signals”. Applied Sciences 8 (9), 1584

Yao, Z., Xie, J., Tian, Y., and Huang, Q. (2019) Using Hampel Identifier to Eliminate Profile-Isolated Outliers in Laser Vision Measurement [online] available from <https://www.hindawi.com/journals/js/2019/3823691/#EEq2> [13 September 2020]

Yazdizadeh, A., Patterson, Z., and Farooq, B. (2019) “Ensemble Convolutional Neural Networks for Mode Inference in Smartphone Travel Survey”. ArXiv:1904.08933 [Cs, Stat] [online] available from <https://arxiv.org/abs/1904.08933> [13 September 2020]

Yousefi, S. and Zohoor, M. (2019) “Effect of Cutting Parameters on the Dimensional Accuracy and Surface Finish in the Hard Turning of MDN250 Steel with Cubic Boron Nitride Tool, for Developing a Knowledged Base Expert System”. International Journal of Mechanical and Materials Engineering 14 (1)

Yu, J., Liang, S., Tang, D., and Liu, H. (2016) “A Weighted Hidden Markov Model Approach for Continuous-State Tool Wear Monitoring and Tool Life Prediction”. The International Journal of Advanced Manufacturing Technology 91 (1–4), 201–211

Yu, Y., Wang, C., Gu, X., and Li, J. (2018) “A Novel Deep Learning-Based Method for Damage Identification of Smart Building Structures”. Structural Health Monitoring 18 (1), 143–163

Yuan, Y., Ma, G., Cheng, C., Zhou, B., Zhao, H., Zhang, H.T., and Ding, H. (2020) “Artificial Intelligent Diagnosis and Monitoring in Manufacturing”. National Science Review [online] 7 (2), 418–429. available from <https://arxiv.org/abs/1901.02057> [13 September 2020]

Zang, H., Liu, L., Sun, L., Cheng, L., Wei, Z., and Sun, G. (2020) “Short-Term Global Horizontal Irradiance Forecasting Based on a Hybrid CNN-LSTM Model with Spatiotemporal Correlations”. Renewable Energy 160, 26–41

Zhang, C., Li, Y., Lin, H., and Yang, B. (2015) “Signal Preserving and Seismic Random Noise Attenuation by Hurst Exponent Based Time–Frequency Peak Filtering”. Geophysical Journal International [online] 203 (2), 901–909. available from <https://academic.oup.com/gji/article/203/2/901/577359> [13 September 2020]

Zhang, C., Yao, X., Zhang, J., and Jin, H. (2016) “Tool Condition Monitoring and Remaining Useful Life Prognostic Based on a Wireless Sensor in Dry Milling Operations”. Sensors 16 (6), 795

Zhang, J., Sun, Y., Guo, L., Gao, H., Hong, X., and Song, H. (2020) “A New Bearing Fault Diagnosis Method Based on Modified Convolutional Neural Networks”. Chinese Journal of Aeronautics [online] 33 (2), 439–447. available from <https://www.sciencedirect.com/science/article/pii/S100093611930278X#b0100> [16 September 2020]

Zhang, J., Deng, C., Zheng, P., Xu, X., and Ma, Z. (2021) “Development of an Edge Computing-Based Cyber-Physical Machine Tool”. Robotics and Computer-Integrated Manufacturing 67, 102042

Zhang, L., Gao, H., Wen, J., Li, S., and Liu, Q. (2017) “A Deep Learning-Based Recognition Method for Degradation Monitoring of Ball Screw with Multi-Sensor Data Fusion”. Microelectronics Reliability [online] 75, 215–222. available from <https://www.sciencedirect.com/science/article/pii/S0026271417300896> [13 September 2020]

Zhang, L., Lin, J., Liu, B., Zhang, Z., Yan, X., and Wei, M. (2019) “A Review on Deep Learning Applications in Prognostics and Health Management”. IEEE Access [online] 7, 162415–162438. available from <https://strathprints.strath.ac.uk/70602/> [13 September 2020]

Zhang, Q.H., Hu, Q., Sun, G., Si, X., and Qin, A. (2013) “Concurrent Fault Diagnosis for Rotating Machinery Based on Vibration Sensors”. International Journal of Distributed Sensor Networks 9 (4), 472675

Zhang, S., Geng, B., Varshney, P., and Rangaswamy, M. (2019) Fusion of Deep Neural Networks for Activity Recognition: A Regular Vine Copula Based Approach [online] available from <https://arxiv.org/pdf/1905.02703.pdf> [13 September 2020]

Zhang, Y.C. and Sakhanenko, L. (2019) “The Naive Bayes Classifier for Functional Data”. Statistics & Probability Letters [online] 152, 137–146. available from <https://www.sciencedirect.com/science/article/pii/S0167715219301208> [13 September 2020]

Zhang, Z., Chen, H., Xu, Y., Zhong, J., Lv, N., and Chen, S. (2015) “Multisensor-Based Real-Time Quality Monitoring by Means of Feature Extraction, Selection and Modeling for Al Alloy in Arc Welding”. Mechanical Systems and Signal Processing [online] 60–61, 151–165. available from <https://www.sciencedirect.com/science/article/pii/S088832701400510X> [13 September 2020]

Zhao, H. and Liu, H. (2019) “Multiple Classifiers Fusion and CNN Feature Extraction for Handwritten Digits Recognition”. Granular Computing

Zhao, R., Wang, J., Yan, R., and Mao, K. (2016) Machine Health Monitoring with LSTM Networks.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., and Gao, R.X. (2019) “Deep Learning and Its Applications to Machine Health Monitoring”. Mechanical Systems and Signal Processing 115, 213–237

Zhao, Z., Lin, P., Shen, L., Zhang, M., and Huang, G.Q. (2020) “IoT Edge Computing-Enabled Collaborative Tracking System for Manufacturing Resources in Industrial Park”. Advanced Engineering Informatics [online] 43, 101044. available from <https://www.sciencedirect.com/science/article/pii/S1474034620300136#f0020> [13 September 2020]

Zhou, C., Guo, K., Yang, B., Wang, H., Sun, J., and Lu, L. (2019) “Singularity Analysis of Cutting Force and Vibration for Tool Condition Monitoring in Milling”. IEEE Access 7, 134113–134124

Zhou, C., Yang, B., Guo, K., Liu, J., Sun, J., Song, G., Zhu, S., Sun, C., and Jiang, Z. (2020) “Vibration Singularity Analysis for Milling Tool Condition Monitoring”. International Journal of Mechanical Sciences [online] 166, 105254. available from <https://www.sciencedirect.com/science/article/pii/S0020740319330954#bib0032> [13 September 2020]

Zhou, L., Li, Q., Huo, G., and Zhou, Y. (2017) Image Classification Using Biomimetic Pattern Recognition with Convolutional Neural Networks Features [online] available from <https://www.hindawi.com/journals/cin/2017/3792805/> [13 September 2020]

Zhou, X. and Wang, J. (2015) “Feature Selection for Image Classification Based on a New Ranking Criterion”. Journal of Computer and Communications 03 (03), 74–79

Zhou, X., Lu, P., Zheng, Z., Tolliver, D., and Keramati, A. (2020) “Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree”. Reliability Engineering & System Safety [online] 200, 106931. available from <https://www.sciencedirect.com/science/article/pii/S0951832019308580#bib0027> [13 September 2020]

Zhou, Y. and Xue, W. (2018) "A Multisensor Fusion Method For Tool Condition Monitoring In Milling". Sensors 18 (11), 3866

Zhou, Z., Huang, G., and Wang, X. (2019) “Ensemble Convolutional Neural Networks for Automatic Fusion Recognition of Multi-platform Radar Emitters”. ETRI Journal 41 (6), 750–759

Zhu, A., Wu, Q., Cui, R., Wang, T., Hang, W., Hua, G., and Snoussi, H. (2020) “Exploring a Rich Spatial–Temporal Dependent Relational Model for Skeleton-Based Action Recognition by Bidirectional LSTM-CNN”. Neurocomputing 414, 90–100

Zhu, J., Chen, H., and Ye, W. (2020) “A Hybrid CNN–LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar”. IEEE Access [online] 8, 24713–24720. available from <https://ieeexplore.ieee.org/document/8978926> [13 September 2020]

Zhu, K., Li, G., and Zhang, Y. (2020) “Big Data Oriented Smart Tool Condition Monitoring System”. IEEE Transactions on Industrial Informatics [online] 16 (6), 4007–4016. available from <https://ieeexplore.ieee.org/document/8918449> [13 September 2020]

Ziegler, S., Woodward, R.C., Iu, H.H.C., and Borle, L.J. (2009) “Current Sensing Techniques: A Review”. IEEE Sensors Journal 9 (4), 354–376

Zong, W., Huang, G.B., and Chen, Y. (2013) “Weighted Extreme Learning Machine for Imbalance Learning”. Neurocomputing 101, 229–242

Zuo, Y. and Drummond, T. (2017) Fast Residual Forests: Rapid Ensemble Learning for Semantic Segmentation [online] available from <http://proceedings.mlr.press/v78/zuo17a.html> [13 September 2020]

Appendix A. Python code

The python code of 1d CNN model:

```
import re
import os
import numpy as np
import pandas as pd
import scipy.io as sio
from pandas import DataFrame as df
from sklearn.model_selection import train_test_split
from numpy import array
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import Flatten
from keras.layers.convolutional import Conv1D
from keras.layers.convolutional import MaxPooling1D
from keras.layers import Conv1D, GlobalAveragePooling1D, MaxPooling1D,Dropout

from keras import Input,Model,optimizers
from sklearn import metrics
from keras.callbacks import Callback

path_save=os.getcwd()+'\\save_files_1d_cnn_paras\\'

data = pd.read_csv("example.csv")
data_new = pd.DataFrame(data)

AC_TF_mat= scipy.io.loadmat('AC_TF_train.mat')
AC_TF_max= AC_TF_mat['AC_TF_train']
pd.DataFrame(AC_TF_max)
target =  AC_TF_max[:,9]

y = df(df(target)[0].map({1:'a',2:'a',3:'a',4:'d'}))
dummy_Y=pd.get_dummies(y.iloc[:,0])

class Metrics(Callback):

    def on_epoch_end(self, epoch, logs = {}):
        self.model.save('{path_save}model_{epoch}.h5'.format(path_save=path_save,epoch=epoch+1))

train_X,test_X,train_y,test_y=train_test_split(X,dummy_Y,test_size=0.3,random_state=1)

model = Sequential()
model.add(Conv1D(filters=8, kernel_size=3, padding='same',activation='relu', input_shape=(23,1)))
model.add(Conv1D(filters=8, kernel_size=3, padding='same',activation='relu'))
model.add(MaxPooling1D(pool_size=2))
model.add(Dropout(0.5))

model.add(Conv1D(filters=16, kernel_size=3, padding='same',activation='relu'))
model.add(Conv1D(filters=16, kernel_size=3, padding='same',activation='relu'))
model.add(MaxPooling1D(pool_size=2))
model.add(Dropout(0.5))

model.add(Flatten())
model.add(Dense(32, activation='relu'))
model.add(Dense(2, activation='softmax'))
model.compile(loss='categorical_crossentropy', optimizer=optimizers.SGD(momentum=0.9,lr=0.003), metrics=['acc
uracy'])

history = model.fit(train_X,train_y,epochs=300,batch_size=8,validation_data=(test_X,test_y),callbacks=[Metric
s()])
```

The python code of CNN+LSTM model:

```
import random
import os
import numpy as np
from sklearn.model_selection import train_test_split
from keras.models import Sequential
from keras.losses import categorical_crossentropy
from keras.optimizers import Adadelta
import pandas as pd
import matplotlib.pyplot as plt
from keras.layers import Conv2D, Conv1D, MaxPool2D, MaxPool1D, Flatten, Activation,Drop
out, Dense, LSTM, SimpleRNN
from keras.optimizers import Adam
from keras.layers import TimeDistributed

height = 300
def get_last_line(filename):
    with open(filename, 'rb+') as f:
#         first_line = f.readline()
        off = -250
        while True:
            f.seek(off, 2)
            lines = f.readlines()
            if len(lines)>=1200:
                return lines[(-1)*height:]
                break
            off *= 2

def read_path(path_name, path_name2, name):
    try:
        labels=np.loadtxt(path_name2,delimiter = ",", skiprows = 1)[:, 1]
    except:
        labels=np.loadtxt(path_name2,delimiter = ",", skiprows = 1)
    data=[]
    for i in range(1, 316):
        full_path = path_name+name+str(i)+'.csv'
        lines=get_last_line(full_path)
        lst=[]
        for line in lines:
            lst.extend(line.decode("utf-8").strip().split(','))
        data.append(lst)
    return data, labels

data, labels=read_path('./save_c1/', 'c1_mean.csv', 'new_c1_')
data2, labels2=read_path('./save_c4/', 'c4_mean.csv', 'new_c4_')
data3, labels3=read_path('./save_c6/', 'c6_mean.csv', 'new_c6_')
```

```
input_file=open('c1c4_all_input_file.csv', 'w')
for i in range(0, 314):
    line=''
    for s in data[i]:
        line=line+s+','
    input_file.write(line+str(labels[i])+'\n')
for i in range(0, 314):
    line=''
    for s in data2[i]:
        line=line+s+','
    input_file.write(line+str(labels2[i])+'\n')

input_file.close()

test_file=open('c1c4_all_test_file.csv', 'w')
for i in range(0, 314):
    line=''
    for s in data3[i]:
        line=line+s+','
    test_file.write(line+str(labels3[i])+'\n')
test_file.close()

input_data=pd.read_csv('c1c4_all_input_file.csv')
train_x = input_data.iloc[:,:-1]
train_y = input_data.iloc[:,-1]
train_x = train_x.as_matrix().reshape(-1,1, height, 7, 1)

test_data=pd.read_csv('c1c4_all_test_file.csv')
test_x = test_data.iloc[:,:-1]
test_y = test_data.iloc[:,-1]
test_x = test_x.as_matrix().reshape(-1, 1,height, 7, 1)

input_shape=[1,height,7,1]
model = Sequential()
model.add(TimeDistributed(Conv2D(8, (3,3), activation='relu'),input_shape=input_shape))
#model.add(TimeDistributed(BatchNormalization()))
#model.add(TimeDistributed(Conv2D(16, (3,3), activation='relu')))
model.add(TimeDistributed(MaxPool2D(pool_size=(2,1))))
#model.add(TimeDistributed(Conv2D(64, (3,3), activation='relu')))
#model.add(TimeDistributed(MaxPool2D(pool_size=(2,1))))
model.add(TimeDistributed(Flatten()))
#model.add(TimeDistributed(Dense(1, activation='linear')))
model.add(Dropout(0.5))

model.add(LSTM(200,return_sequences=False))##size can not too low,200 is proper
model.add(Dropout(0.5))
model.add(Dense(200, activation='linear'))
model.add(Dense(1))
model.compile(loss='mae', optimizer='adam')
model.fit(train_x, train_y, batch_size=32, epochs=300, validation_data=(test_x, test_y
), verbose=2, shuffle=False)
```

The python code of 2D CNN model:

```
from sklearn.preprocessing import LabelBinarizer
from sklearn.preprocessing import MinMaxScaler
import pandas as pd
import numpy as np
import glob
import cv2
import os
import matplotlib.pyplot as plt
from keras.models import Sequential
from keras.layers import Conv2D, MaxPool2D, Flatten, Dropout, Dense,AveragePooling2D
from keras.losses import categorical_crossentropy
from keras.optimizers import Adadelta
from keras_preprocessing.image import ImageDataGenerator

df_AC_train=pd.read_csv('JLD_case1&3_p_train.csv',header=None,names=['id','wear'])
df_AC_val=pd.read_csv('JLD_case2_p_val.csv',header=None,names=['id','wear'])
#######
df_BC_train=pd.read_csv('JLD_case2&3_p_train.csv',header=None,names=['id','wear'])
df_BC_val=pd.read_csv('JLD_case1_p_val.csv',header=None,names=['id','wear'])
########
df_AB_train=pd.read_csv('JLD_case1&2_p_train.csv',header=None,names=['id','wear'])
df_AB_val=pd.read_csv('JLD_case3_p_val.csv',header=None,names=['id','wear'])

datagen=ImageDataGenerator(rescale=1./255)
train_generator_AC=datagen.flow_from_dataframe(dataframe=df_AC_train,directory="JLD_case1&3_p_train",
                                            x_col="id", y_col="wear",
                                            class_mode="raw",  shuffle=False,validate_filenames=True,target_s
ize=(28,28), batch_size=32)

valid_generator_AC = datagen.flow_from_dataframe(dataframe=df_AC_val,directory="JLD_case2_p_val",x_col="id",
y_col="wear",
                                              class_mode="raw",shuffle=False,validate_filenames=True,target_s
ize=(28,28),batch_size=32)

def define_model():
    model = Sequential()
    model.add(Conv2D(32, (3, 3), activation='relu', kernel_initializer='he_uniform', padding='same', input_sh
ape=(28, 28, 3)))
    #model.add(Conv2D(32, (3, 3), activation='relu', kernel_initializer='he_uniform', padding='same'))
    model.add(AveragePooling2D((2, 2)))
    model.add(Dropout(0.2))
    model.add(Conv2D(64, (3, 3), activation='relu', kernel_initializer='he_uniform', padding='same'))
    model.add(AveragePooling2D((2, 2)))
    model.add(Dropout(0.3))

    model.add(Flatten())
    model.add(Dense(120, activation='relu', kernel_initializer='he_uniform'))
    model.add(Dense(84, activation='relu', kernel_initializer='he_uniform'))
    model.add(Dense(1, activation='linear'))
    model.compile(optimizer='adam', loss='mse', metrics=['accuracy'])
    return model

STEP_SIZE_TRAIN_AC=train_generator_AC.n//train_generator_AC.batch_size
STEP_SIZE_VALID_AC=valid_generator_AC.n//valid_generator_AC.batch_size

history=model.fit_generator(generator=train_generator_AC, steps_per_epoch=STEP_SIZE_TRAIN_AC,
                            validation_data=valid_generator_AC, validation_steps=STEP_SIZE_VALID_AC, epochs
=50, verbose=1)
```

Appendix B. Ethical forms

Certificate of Ethical Approval

Applicant:

XiaoYang Zhang

Project Title:

On-line prediction of tool life of CNC machine using multi-sensor monitoring system

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Low Risk

Date of approval:

29 June 2017

Project Reference Number:

P53201

Certificate of Ethical Approval

Applicant:

XiaoYang Zhang

Project Title:

A multi-sensor based online tool condition monitoring system for milling process

This is to certify that the above named applicant has completed the Coventry University Ethical Approval process and their project has been confirmed and approved as Low Risk

Date of approval:

23 July 2018

Project Reference Number:

P72626